DTIC FILE COPY

(2)

AFHRL-TR-89-56

AIR FORCE

HUMAN RESOURCES LABORATORY

AD-A221 004

AIR FORCE OFFICER QUALIFYING TEST (AFOQT) FORM P: TEST MANUAL

Frances R. Berger
Willa B. Gupta
Raymond M. Berger

Psychometrics Inc.
13245 Riverside Drive
Sherman Oaks, California 91423-2172

Jacobina Skinner

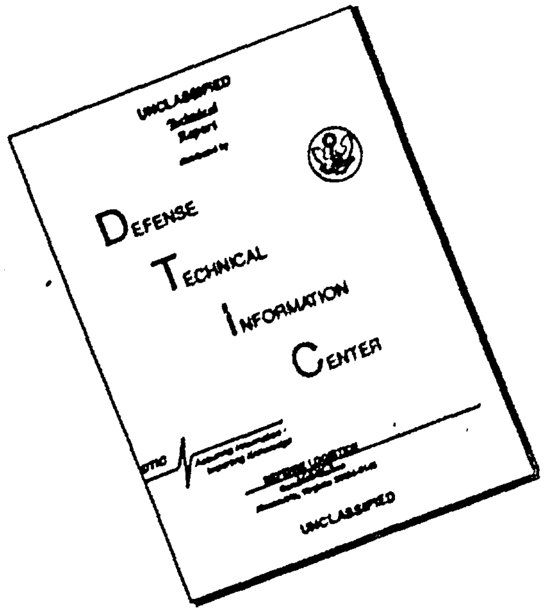
MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

April 1990
Interim Technical Report for Period January 1987 - September 1989

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

NOTICE

This report has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1990	Interim - January 1987 to September 1989

4. TITLE AND SUBTITLE

Air Force Officer Qualifying Test (AFOQT)
Form P: Test Manual

5. FUNDING NUMBERS

C - F33615-83-C-0035
PE - 62703F
PR - 7719
TA - 18
WU - 24

6. AUTHOR(S)

Frances R. Berger Raymond M. Berger
Willa B. Gupta Jacobina Skinner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Psychometrics Inc.
13245 Riverside Drive
Sherman Oaks, California 91423-2172

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFHRL-TR-89-56

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The latest form of the Air Force Officer Qualifying Test (AFOQT)--Form P--is the first AFOQT to be developed in two parallel forms, Forms P1 and P2. The AFOQT has been part of the selection process for officer commissioning programs and for specific programs such as pilot and navigator training since 1951. Forms P1 and P2 are organized into 16 subtests which form five composites: Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative.

The present report is a test manual for Form P. It describes the development of Form P; its continuity with Form O via common items and characteristics; statistical data for Forms P1 and P2, such as reliability, item difficulty, item discrimination, score characteristics, and norms; test validity derived from studies of other AFOQT forms; and considerations of administration, scoring, and interpretation. Sample questions and pertinent bibliographic references are included.

14. SUBJECT TERMS

Air Force Officer Qualifying Test
aptitude tests
mental abilities testing
officer classification
officer selection
test manual

15. NUMBER OF PAGES

110

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

SUMMARY

This test manual provides information on the use and interpretation of scores derived from the Air Force Officer Qualifying Test (AFOQT) Form P. Since 1951, the AFOQT has been part of the selection process for officer commissioning programs and pilot and navigator training. The latest of 16 successive forms of the AFOQT, Form P, is the first to be developed in two parallel forms.

The manual briefly describes the evolution of the AFOQT across its periodic revisions designed to improve predictive validity and prevent test compromise. A description of each subtest and composite is provided, and sample problems are presented.

Experimental AFOQT items were constructed and field tested with the goal of providing a large AFOQT item bank. Analyses of the data provided information important for Form P item selection. Form P, like Form O, contains 380 items organized into 16 subtests which form five composites: Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative. Resemblance to Form O in terms of psychometric properties, content, and style was the criterion for Form P item selection. Preliminary analyses of Forms P1 and P2 based on the test data of airmen attending Basic Military Training at Lackland AFB established that the forms are highly similar to each other, as well as being comparable to Form O.

Additional evidence of the comparability of the forms is given in some detail for officer applicant samples tested during the Initial Operational Test and Evaluation (IOT&E) of AFOQT Form P. Central characteristics, reliability (KR-20), standard error of measurement, item difficulty and discrimination, and item content are reported. Intercorrelations among the Form P subtests are interpreted for construct validity. Included are summaries of criterion-related validity studies relating AFOQT Form O to performance measures for officer commissioning programs at Officer Training School (OTS) and the Air Force Reserve Officer Training Corps (AFROTC); entry-level pilot and navigator training; and non-rated officer training. Statistically significant positive validity coefficients were obtained in all of the studies. Validity studies for Form P are planned, and it is expected that the predictive validity of Form P will be similar to that for Form O because of the comparability of Forms O and P.

Test administration and scoring considerations are discussed. Topics include testing locations, standardized procedures, test security, testing schedule, and score interpretation. Research on a Quick Score procedure to prescreen applicants for OTS and AFROTC is reviewed. Appropriate documents for the use and interpretation of the AFOQT are cited.

PREFACE

The Air Force Human Resources Laboratory (AFHRL) is designated as the primary test development agency for the Air Force Officer Qualifying Test (AFOQT) by Air Force Regulation 35-8, Air Force Military Personnel Testing System. The current research and development (R&D) effort was undertaken as part of AFHRL's responsibility to develop, revise, and conduct research in support of the AFOQT. Work was accomplished under Task 771918, Selection and Classification Technologies, which is part of a larger effort in Force Acquisition and Distribution Systems. The project was completed under Work Unit 77191824, Officer Item Pool Development, as a contractual effort (Contract F33615-83-C-0035) by Psychometrics Inc., Sherman Oaks, CA. Ms Jacobina Skinner represented the monitoring agency as contract manager.

TABLE OF CONTENTS

LIST OF TABLES

AIR FORCE OFFICER QUALIFYING TEST (AFOQT) FORM P:

TEST MANUAL

I. BACKGROUND

Since 1951, when the test appeared in a preliminary version, the Air Force Officer Qualifying Test (AFOQT) has been part of the selection process for Air Force officer commissioning programs and for pilot and navigator training programs. The latest of 16 successive forms of the test, Form P, was the first to be constructed in two parallel forms. The development, measurement characteristics, and equivalence of these forms, and the use and interpretation of their scores, are the principal topics of this manual.

History of the AFOQT

Purpose

As with the other military services, the U.S. Air Force has the goal of commissioning officers who will be leaders with exceptional levels of ability. The primary selection criteria for officers are education and aptitude (Brown, 1987). The AFOQT is used to measure the aptitudes of candidates for two of the three officer commissioning programs: the Air Force Reserve Officer Training Corps (AFROTC) and the Officer Training School (OTS). (The third program, which is offered at the Air Force Academy in Colorado Springs, Colorado, uses other criteria for selection and is exempt from the AFOQT requirement.) Aircrew classification, a function that matches aptitudes to the task requirements of pilots and navigators, constitutes the second major use of the AFOQT.

AFOQT Form P contains 380 items organized into 16 subtests which form five composites: Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative (Table 1). A candidate for admission to AFROTC, OTS, Undergraduate Pilot Training (UPT), or Undergraduate Navigator Training (UNT) must meet or exceed minimum qualifying scores on the AFOQT composites to continue in the selection process. The validity of the procedure has undergone continual study and verification to provide assurance that the AFOQT is a highly suitable instrument for officer selection and classification.

Evolution of the AFOQT

The information in this section is drawn largely from Rogers, Roach, and Short (1986), who noted that age, education, and medical fitness requirements were the principal considerations for selection to pilot training programs before World War II. Early efforts to use physiological or psychological tests to select student pilots were only partially successful. Following the onset of U.S. participation in World War II, the Aviation Psychology Program was initiated, with the mission of developing pilot selection instruments. Among the prominent psychologists and measurement specialists staffing this program were John C. Flanagan, J. P. Guilford, Robert L. Thorndike, and Neal E. Miller. The need for a large number of aircrew personnel, as well as the shortage of physically qualified

Table 1. Organization of AFOQT Form P Subtests and Composites

Subtest	No. of items	Composite				
		Pilot	Nav-Tech	Academic aptitude	Verbal	Quantitative
Verbal Analogies	25	X		X	X	
Arithmetic Reasoning	25		X	X		X
Reading Comprehension	25			X	X	
Data Interpretation	25		X	X		X
Word Knowledge	25			X	X	
Math Knowledge	25		X	X		X
Mechanical Comprehension	20	X	X			
Electrical Maze	20	X	X			
Scale Reading	40	X	X			
Instrument Comprehension	20	X				
Block Counting	20	X	X			
Table Reading	40	X	X			
Aviation Information	20	X				
Rotated Blocks	15		X			
General Science	20		X			
Hidden Figures	15		X			
Total	380					

college students in 1941, dictated their first task--to develop a general abilities test to replace the previously used college requirement. The result was the Aviation Cadet Qualifying Examination (ACQE).

Initial Screening Tests. The ACQE, a general intelligence test consisting of 150 items, was taken by more than a million men during the war to qualify as commissioned officers and pilots, bombardiers, and navigators.

The title of this screening test was changed to the Army Air Forces Qualifying Examination (AAFQE) in June 1944, due to the fact that the test was by then being administered to enlisted men (to serve as gunners) as well as to aviation cadets.

Selection and Classification Tests. The need for a test to screen and classify potential aircrew members resulted in the development of the Aircrew Classification Battery (ACB). This battery included psychomotor tests and both power and speeded paper-and-pencil tests that provided Pilot, Navigator, and Bombardier stanines (composite scores). The ACB was revised many times and was eventually used for selection as well as classification.

The ACB was discontinued in 1947, because of a decrease in applicant flow, but was reinstated in 1951 at the request of Headquarters Air Training Command, which was finding it difficult to find aviation cadet applicants who had 2 years of college. In the interim, the Aviation-Cadet Officer-Candidate Qualifying Test (AC-OC-QT) was used experimentally to screen applicants for Officer Candidate School (OCS) (now OTS) and for direct commissioning, to

screen aviation cadet applicants, and to measure necessary nonflying aptitudes such as electronics. The AC-OC-QT became operational in October 1950 under the name Aviation Cadet Qualifying Test (ACQT). The ACQT, because of its ability to predict success on the ACB, replaced the ACQE as a screening device.

The AFOQT. The first version of the AFOQT, which incorporated the AC-OC-QT, was designed in 1951 to predict success in OCS and to screen for aircrew training. Form A, which consisted of the USAF Officer Activity Inventory, the Attitude Survey, and the preliminary AFOQT, followed. Five composite scores were derived from Form A: Officer Quality, Observer-Technical, Pilot, Verbal, and Quantitative. Form B, designated the Officer Selection Form, became operational in September 1955, replacing the ACB for aircrew selection. The AFOQT forms that followed (Forms C through L) underwent frequent revisions to composites, subtest content areas, test items, conversion tables, and norms.

Precursors of Form P. The history of the development of Forms M, N, and O shows the way in which the design of Form P evolved. Appendix A displays the content and organization of these forms. Form M (1975) (Table A-1) was the same as its predecessor, Form L, with the exception that it provided a separate female-specific conversion table for the Pilot composite (Miller, 1974). A comparison of Tables A-1 and A-2 indicates the structural changes introduced into Form N. The need for revision of Form M was revealed by a validity study conducted by Valentine (1977). The five booklets of Form M were reduced to four. Five subtests were removed: Officer Biographical Inventory, Pilot Biographical Inventory, Aviation Information, Visualization of Maneuvers, and Stick and Rudder Orientation. Seven subtests were added: Background for Current Events, Pilot Biographic and Attitude Scale, Table Reading, Electrical Maze, Block Counting, Tools, and Rotated Blocks. Further, the Quantitative Aptitude and Verbal Aptitude composites in Booklets 1 and 2 were subdivided into subtests. Three subtests--Arithmetic Reasoning, Math Knowledge, and Data Interpretation--replaced Quantitative Aptitude. Word Knowledge, Reading Comprehension, Background for Current Events, and Verbal Analogies replaced Verbal Aptitude.

Changes were also made to the normative base for Form N (Gould, 1978). The norming sample, drawn from the three commissioning sources was well as active duty second lieutenants, was designed to reflect the range of ability expected in the officer applicant population. Norms were developed for the total group and for three educational-level groups: less than 2 years of college, more than 2 years (but not graduates), and college graduates. New conversion tables were developed for the three educational-level groups.

Form O (1981) dropped the subtests titled Background for Current Events, Tools, Aerial Landmarks, and Pilot Biographic and Attitude Scale, and added Aviation Information and Hidden Figures (Rogers, Roach, & Wegner, 1986). Form O was the first AFOQT form to be equated to an anchor test, Form N, through items common to both forms. A further change involved the number of conversion tables, which were reduced to five--one for each composite--as it was no longer considered relevant to emphasize educational differences in the conversions (Roach, 1986; Roach & Rogers, 1982). Form O was also associated with administrative changes that promoted efficiency in test administration

and reduced technical problems in achieving scoring accuracy. That is, all subtests were enclosed in a single booklet; all answer sheets were electronically scanned and computer-scored; and there were two central scoring locations, Maxwell AFB, Alabama, for AFROTC test administrations, and Brooks AFB, Texas, for all others.

One of the most salient features of the later AFOQT forms is their continuity, maintained by the use of "anchor" or common items. Form P, like Form O, has a certain proportion of items taken from the previous version. The essential nature of the aptitude measures found by the Air Force to be related to proficient officer performance remained unchanged in AFOQT Form P.

Description of the Subtests

"Aptitude" is usually interpreted to mean inherent or natural ability and "aptitude test," a measure of the ability to learn. In practice, aptitude tests frequently measure combinations of potential abilities and learned skills. The Math Knowledge and Arithmetic Reasoning tests in the AFOQT, for example, resemble achievement tests and assume some prior knowledge. The AFOQT can be seen as an aptitude test battery, however, in that the prediction of future performance--both in training and on the job--is its primary purpose. In addition, the AFOQT serves a classification function in the Air Force personnel system by distinguishing between such assignments as pilot training and navigator training (Miller, 1969a). The AFOQT subtests and composites, and the abilities they measure, are described below. Sample problems, taken from the Air Force Officer Qualifying Test Information Pamphlet (Department of the Air Force, 1987) are provided in Appendix B. The sample problems are provided for illustrative purposes only and should not be viewed in and of themselves as a blueprint for the test. They reflect the difficulty and content of AFOQT items but do not necessarily conform to all of the subtest specifications in the descriptions which follow.

Verbal Analogies (VA)

The items in this subtest take two forms. The first is a pair of words joined by the phrase "is to" to express a relationship between them, followed by a third word, the stimulus. The examinee must choose from a list of five words the word that has the same relationship to the stimulus as that between the given pair. The second form displays in the stem only a pair of words with an implied relationship. Among five alternative word pairs is one that expresses the same type of relationship as that of the given pair.

The relationships between words are of several types, including cause and effect, sequence, antonym, synonym, degree, action to object, object to action, part to part, part to whole, member to category, and object to attribute. Among the types of wrong alternatives may be homonyms or antonyms of the correct answer; wrong tenses or grammatical forms; or reversals of the direction of the given analogy. To preclude ambiguity, the alternatives are constructed such that only one type of relationship in a given word pair can be inferred.

(An example of ambiguity: Morning is to evening as

A. breakfast is to dinner
B. sunlight is to starlight

(Either alternative would be an appropriate analogy.)

Verbal ability (i.e., proper use of vocabulary in various contexts) is one component of the ability required by the VA test. Inferring relationships is another component. The degree of remoteness of these relationships contributes to the difficulty levels of the VA items. The more inferences required, the more difficult the item. For example, the relationship "grass:green" requires the examinee to make fewer inferences than "grass:landscape." The less central the attributes of the analogy are to its semantic representation, the more difficult the item. For example, the relationship "grass:green" is probably more central to our conceptual representation of grass than the relationship "grass:landscape." In general, the "X:Y::A:___" format is easier than the "X:Y::___:___" format, in that the former directs the search for alternatives.

VA items do not test knowledge of specific subject matters. The vocabulary used is neither complex nor esoteric.

Arithmetic Reasoning (AR)

The AR items test the ability to understand and manipulate relationships in order to arrive at solutions to problems. The major emphasis is on reasoning, but the problems require knowledge of basic arithmetic functions such as addition, subtraction, multiplication, division, percentages, ratios, proportions, squares, square roots, or the conversion of dimensions in a single measurement system (e.g., yards to feet). Knowledge of the formulas used to calculate the lengths of the sides of right triangles or the circumferences and areas of circles may be necessary to solve some problems. In order to avoid irrelevant sources of difficulty, calculations are simple and not time-consuming.

The AR items are constructed as word problems. In general, a characteristic such as height, weight, speed, distance travelled, temperature, or interest earned must be calculated in the context of other characteristics that have a given relationship (e.g., half as fast) to a given object. Calculations may involve the conversions of dimensions in a scale, a requirement implied by the problem. For example, if told that a line is 20 yards long and asked how many 3-foot-long objects fit on that line, the examinee must infer that conversion from yards to feet is necessary. Verbal ability as an intervening variable is minimized by use of basic, simple vocabulary in the item stems.

Knowledge of specific subjects is not necessary for the AR items. Measured by this subtest are the abilities to analyze problems and use reasoning to manipulate the components analyzed, in order to satisfy the problem specifications.

Reading Comprehension (RC)

RC items require the examinee to read and understand paragraphs. Single reading passages ranging from approximately 40 to 150 words in length are followed by five alternatives each, one of which completes the final sentence of the passage appropriately. The incorrect alternatives may complete the last sentence in a grammatical sense, but they show either an incomplete or incorrect understanding of the given passage. The subject matters of the reading passages include art, economics, the environment, literature, philosophy, psychology, science, and sociology. Knowledge of these subjects is not necessary for the examinee to arrive at correct answers. Wrong alternatives are typically ones that refer to elements of the passage but fail to capture the central theme or continue the logic.

The comprehension required by this subtest depends upon one's abilities to abstract, generalize, synthesize, recognize concepts, and reason constructively. The relevance of this subtest clearly lies in its relationship to the informational aspects of high-level Air Force jobs.

Data Interpretation (DI)

The DI subtest consists of data presented in tables or graphs. These data are frequently expressed as units (such as thousands of hours) which are identified either above the data in a column heading or below the data in a footnote. Each set of data is followed by several questions pertaining to it, and each question has five alternatives. The answer to any one item (question) is independent of the answers to the other items that refer to the same data.

All of the DI questions require examination and interpretation of the relationships within the data displayed. Answers to some questions depend on making the correct associations between two or more sets of numeric information given in different areas of a table. Other questions require a series of simple arithmetic calculations or estimates such as the computation of ratios, percentages, proportions, or ranges after interpreting the given data. Most of these calculations involve whole numbers, and the alternative answers rarely have more than one decimal place.

Incorrect alternatives may be values or variables obtained by locating and/or comparing incorrect cells, arriving at intermediate solutions, performing incorrect computations, misunderstanding the meaning of cells, and/or failing to convert cell numbers to their correct scales. The subject matter of the charts does not play a significant role in the examinee's ability to answer the DI questions.

The DI subtest calls upon one's abilities to locate, interpret, and integrate different types of data appearing in a variety of tabular and graphic structures, in order to satisfy problem requirements. Recognizing relationships, as well as basic arithmetic reasoning, are other abilities involved in this test.

Word Knowledge (WK)

The WK subtest measures vocabulary acquisition, a straightforward aspect of verbal ability. The items are cast as synonyms. A word is given and then followed by five alternatives; one of these alternatives has the same meaning as the given word. Incorrect alternatives include words that may sound or look like the given word, have a synonym that sounds or looks like the given word, are opposite in meaning, or are a commonly mistaken meaning for the word. No technical or scientific terms, foreign expressions, or proper names are included. All words are in the singular form unless they have acquired a use more common in plural form (e.g., data, cross-purposes). The vocabulary is at the high school completion level.

Math Knowledge (MK)

The MK subtest presents items requiring knowledge of mathematical relationships. The item stems are either word or number problems. The item types include symbolic fractions, decimals, and square root solutions. Understanding of mathematical terms such as "reciprocal" or "prime number" is assumed by some items. Incorrect answers to the problems may reflect solutions arrived at because of common misunderstandings, lack of knowledge, or inability to apply learned mathematical relationships accurately.

Items are worded such that reading ability does not play a significant role and none of the calculations involved is time-consuming, complex, or difficult. MK measures mathematical reasoning and mathematical knowledge typical at a first-year college level.

Mechanical Comprehension (MC)

The MC subtest requires acquaintance with and understanding of basic physical principles such as friction, centrifugal force, and pressure, and the ways in which these principles apply to the operation of mechanical devices. Diagrams of mechanical situations are associated with 75% of the items; each diagram is followed by several items. These items measure such things as the understanding of transfer of rotational motion and rotational range of action. The remaining items, unillustrated, require knowledge of the mechanical parts of engines (particularly automobile engines) and knowledge of hardware and tools. Both diagrammatic and unillustrated questions deal with functions of parts, the effect of movement of one or more parts on other parts, and the actions necessary from certain parts to achieve a designated effect. The incorrect alternatives to the MC questions may be in terms of incorrect directions (e.g., left rather than right), incorrect measurements, or incorrect application of physical principles.

The mechanical devices or situations in the items are relatively frequently encountered and do not in and of themselves require special knowledge. An individual who achieves a high score on MC is one who has a ready understanding of the principles behind the design of mechanisms and their operation.

Electrical Maze (EM)

EM items ask the examinee to choose a correct path through a maze from among five choices. In each item, five boxes of equal size are lined up at the bottom of a maze. Within each box are two dots marked S (starting point) and F (finishing point). A line leads from each of the dots marked S. Only one box will have a line from the S point, which leads to a single circle at the top of the maze and back to the F point in the same box. Black dots which appear where two lines meet or intersect show the only places where turns or directional changes may be made. If lines meet or cross where there is no dot, turns or directional changes may not be made. There are no continuous loops through the circle. Lines are wrong if they lead to a dead end, if they come back to the box without going through the circle, or if they lead to other boxes. Easy items tend to have fewer lines and dots than the more difficult items.

Solvers of EM problems tend to take a trial-and-error approach. The examinees with the highest scores may develop strategies to eliminate wrong lines quickly. Spatial aptitude in the sense of being able to ignore confusing visual cues on the way to a goal is what is being measured. With respect to aviation, an obvious application of this ability is in tasks involving electrical wiring, but this ability may also be related to piloting tasks involving visually seeking locational cues.

Scale Reading (SR)

The SR subtest necessitates reading gauges, dials, and meters. Drawings of these devices are associated with either one, two, or three items. Each device contains one or two scales which may be straight or curved, and which are divided and subdivided by short vertical or horizontal lines. The width of subdivisions may be of equal intervals or based on logarithmic scales. Numeric values are given at two or more points on each scale. The examinee's task is to estimate the numeric value indicated by an arrow pointing to the outside edge of the scale at a point which is either between or outside the two given values. The arrow is accompanied by the item number so that the examinee can identify the appropriate set of five alternative choices before selecting the alternative that is closest to his or her estimate.

The more difficult items employ straight-line logarithmic scales, curved scales, scales reading from right to left, or combinations of these. The SR subtest measures visual-perceptual acuity, ability to comprehend different types of scales and unit relationships, and ability to count in different scale units.

Instrument Comprehension (IC)

IC items require the examinee to read airplane instruments to determine the position of an airplane in flight. No previous experience with reading airplane instruments is needed. Full explanations and sample items are given at the beginning of the test. Each item depicts two aircraft instruments. One indicates the amount of climb or dive, and the degree of bank to the left or right. The other is a compass that shows the direction in which the airplane is headed. The two instruments are accompanied by four drawings of

aircraft in flight. The examinee must select the one illustration that corresponds to the readings on the instruments.

IC items vary in difficulty depending on the extent to which the amount of climb or dive, degree and direction of bank, and compass heading are varied. Easier items, for example, have one or more of the following elements: no banking, no climb or dive, and a heading in the direction of one of the major compass points. The IC items draw upon not only form visualization and scale but also their relationship, a more complex aptitude.

Block Counting (BC)

The task in the BC subtest is to determine the number of blocks that are touched by a designated block in an illustrated pile. Each pile consists of 11 blocks. At least one full edge of every block in the pile is visible. The pile is not a perfect cube; the twelfth block is "missing" in order to provide visual information on the other blocks. The location and size of each side of the missing block (usually the top level) are clear to the viewer, and two of its six surfaces are obvious from the gap in the cube.

Each block-pile illustration serves five items. Five blocks in each pile are assigned numbers corresponding to five item numbers. The examinee must count how many blocks in the pile touch the numbered block and choose the one answer of five alternatives that matches the count.

The ability to visualize the abutment of surfaces to designated objects is what is measured by this test. It is one of several distinct approaches in the AFOQT used to assess spatial ability.

Table Reading (TR)

The TR subtest is a speeded test; that is, it contains more items than can usually be answered correctly in the allotted time. It tests one's ability to read tables quickly and accurately to locate and identify the intersection of two coordinates. Examinees are asked to locate a number in the cell of a table, given the X and Y coordinates of the cell. Examinees select their answer from one of five alternatives.

The table is a square matrix containing many cells. The column and row headings are the same and range from a negative number to its equivalent positive number. Numbers inside the cells increase from left to right and from top to bottom. A cell number may be identical to the cell number in the cell above or to the left of it or may be several units larger. Given particular X and Y coordinates, the incorrect alternatives can be the cells found by applying incorrect combinations of signs to the coordinates, by using numbers in cells adjacent to the correct cell, or by using numbers that contain the same digits as those in the correct cell, but in a different order. The easier items are those in which the correct cells are on or near the border of the table and where the X and Y coordinates are of the same sign.

Aviation Information (AI)

AI items measure knowledge of general aeronautical concepts and terminology. Examinees must choose the one of five alternatives that either best completes a sentence or best answers a question. Items cover topics

such as the functions of certain airplane parts, the proper operation of airplanes, hazards encountered by airplanes and how these may differ from one type of airplane to another, navigation and communication techniques, aviation weather maps, and aeronautical terms and concepts. AI items are relevant to civil aeronautics rather than military aeronautics. The emphasis of the items is on light aircraft or general aviation aircraft with turbo jet or reciprocating engines.

Rotated Blocks (RB)

The RB subtest requires the mental manipulation of objects in three-dimensional space and visualization of the changed appearance of those objects as they are rotated. Each item presents a drawing of a given three-dimensional block followed by five drawings depicting blocks in varied positions. One of the five has the exact configuration of the given block but is shown from a different angle. The examinee must determine which one of the five has this configuration. The drawings are made large and clear to preclude calling upon acuity of visual discrimination.

The aptitude measured by RB is spatial visualization.

General Science (GS)

GS items measure knowledge of basic scientific terms, concepts, principles, and instruments. Questions are followed by five alternative choices. The content is drawn from astronomy, biology, chemistry, electronics, geography, meteorology, and physics.

Hidden Figures (HF)

HF items examine an individual's ability to see a simple figure in a complex drawing. Five distinct figures lettered A, B, C, D, and E appear at the top of each page. Below these are several drawings, each embedding one of the five distinct figures in a complex setting. The examinee must determine which one of the five figures appears in each drawing. The correct hidden figure in each complex drawing will always be the same size and in the same position (unrotated) as it appears among the figures at the top of the page.

HF calls upon visual perception, spatial imagery, short-term memory, and the ability to ignore irrelevant and confusing visual cues.

Composites

The subtests that make up the composites were shown in Table 1. Logical analysis and empirical data, where possible, have determined the recommended uses of these composites (Miller, 1969b).

Pilot (P)

The Pilot composite was formulated to predict Undergraduate Pilot Training success. The criterion measure used to evaluate the predictiveness of this composite was flying deficiency leading to removal from training.

High scorers have been shown to have the degree and kinds of aptitudes necessary to complete training successfully.

Navigator-Technical (N-T)

The N-T composite is relevant for success in Undergraduate Navigator Training, other training programs which stress mechanical and engineering concepts, and in pilot training as well.

Academic Aptitude (AA)

The AA composite, which emphasizes verbal and mathematical reasoning skills, was designed to predict success in any training program that is high in academic content. The academic portions of the OTS and AFROTC program curricula are examples.

Verbal (V)

The Verbal composite was formulated to predict training success in programs that require verbal skills, such as administrative services, public information, and education and training.

Quantitative (Q)

The Q composite predicts success in mathematically referenced training programs such as accounting, auditing, statistical services, disbursing, and supply.

II. DEVELOPMENT OF FORM P

In 1983, anticipation of the need for future forms of the AFOQT prompted initiation of a project to develop a pool of experimental items from which selections would be made for AFOQT Form P in two parallel versions. The advantages of constructing two forms concurrently were expected to include improved retesting capability and reduction of opportunities for test compromise.

Approximately 4,800 items were created for the experimental pool. The major requirement guiding item development was that the subtests were to have a high resemblance to the Form O subtests in content, appearance, and distributions of item difficulty and item discrimination. The continuity of instruments proven to predict officer performance would thus be assured. The process of Form P construction summarized in this section of the manual is described at length in Berger, Gupta, Berger, and Skinner (1988).

Experimental Item Construction

New Items

The experimental items were written to be consistent with Form O in terms of format, semantics, and punctuation; content categories of certain subtests; item difficulty range; and specific requirements for graphics. Item writers were provided with operational definitions that described the

constraints they must observe as to the content, scope, complexity, length, appearance, and number of items and response options (see Appendix B, Sample Items). When appropriate, the content categories covered in selected subtests were expanded to include a broader concept of the area tested; for example, adding computer-related questions to the General Science subtest because of their importance in the science fields. Form O item data from officer applicant examinees were analyzed to gain insight on the factors affecting difficulty, and these factors were used to guide Form P item development to approximate Form O's range of difficulty.

Common Items

Included in the research of the experimental pool were anchor or common items selected from Form O. Seven booklets of new and common items were assembled for each subtest except Scale Reading and Table Reading (speeded tests); for these two subtests, 14 booklets were prepared. (To obtain accurate item statistics for the last items in the speeded tests, the tests were administered in either forward or reverse order to seven samples each.) The common items for each subtest appeared in all seven (or 14) booklets of that subtest, in order to verify that the different samples tested on the seven booklets of a subtest were comparable in ranges of ability. Comparability was essential to maximize the constancy of scale indices for item difficulty and item discrimination measures across examinee groups. Based on the collective set of findings described below, the samples were judged to be sufficiently comparable to proceed with a Form P test construction strategy in which item indices were treated as non-sample-specific. A second important reason for including common items was to provide continuity with previous AFOQT forms.

Field Tests

Approximately 350 airmen attending Basic Military Training at Lackland AFB, Texas, were tested on each of the experimental booklets between August 1984 and December 1986. Basic airmen constituted the only practicable group on which to obtain preliminary data for evaluating item adequacy. Because the AFOQT is administered for operational selection and classification purposes at about 500 military testing sites in the Continental United States and overseas, it was not logistically or economically feasible to field test the several thousand new test items with officer applicants. Supplemental data on two subtests, General Science and Aviation Information, were obtained by readministering the experimental test booklets to cadets attending OTS between October 1985 and January 1986. Biographic information was collected from all examinees.

Analysis and Results

Sample Comparability

To evaluate the comparability of the samples of test-takers, the 7 to 14 basic airmen samples tested on each experimental subtest were compared for biographic and score characteristics. Tables which show the analysis results are presented in Berger et al. (1988). A summary of the major findings follows. The samples appear to have been relatively comparable on most

variables. Age differences across samples for each subtest were typically less than 1 year; the mean number of years of education rarely varied by more than 2 months; and the range of differences in percentage of ethnic representation was rarely more than 10%. However, the range of male/female representation in the various samples did vary considerably for some subtests.

The various airmen samples were also compared for test performance, first on the Armed Forces Qualification Test (AFQT) composite score derived from the Armed Services Vocational Aptitude Battery (ASVAB). The mean percentile scores were reasonably close; differences across the samples within subtests ranged from 2.2 (BC) to 6.2 (AR). The modal difference was approximately three points. Performance on the common items included in the experimental AFOQT test booklets provided a second method for assessing the comparability of the airmen samples. On the average, the proportion of airmen samples answering the items correctly differed by .09 or less.

The OTS cadets who took the AI and GS tests were older than the majority of airmen examinees (mean ages 25 to 26 years and 19 to 20 years, respectively). All cadets had completed at least 16 years of education, whereas 95% of the airmen had completed 12 to 14 years. The gender representation was relatively stable across samples within subtests, with the percentage of males varying between 85% and 90%. Ethnic representation among cadets showed a pattern quite similar to that of airmen, but there was a slightly higher proportion of White cadets. On the common items in the experimental test booklets, cadet samples performed more consistently than did the airmen samples. Mean scores for the cadets differed by only about .02 points across the seven samples administered both the Aviation Information and General Science subtests.

Test Statistics

In view of the finding that samples administered each subtest were comparable on biographic and ability indicators, the experimental items were evaluated by classical item analysis procedures within each sample. These analyses were used to identify candidate items for a reduced pool from which Form P selections would be made.

Item Analyses. The results of analyzing the experimental items indicated that the number of new items that met discrimination index and difficulty standards was sufficient to proceed with the development of Form P. Requirements were that item-total score biserial correlations be negative for all non-keyed (incorrect) alternatives and positive for the keyed (correct) responses. Further, biserial correlations for keyed responses had to reach or exceed a minimal discriminative standard of .40. An additional consideration was the need for items to have difficulty levels that fell within the range of and approximated the distribution of difficulty for Form O items.

Parallel Forms Development

The criteria for selecting both common and new items for Forms P1 and P2 from the experimental pool included equivalent mean difficulty levels for each subtest, comparable mean biserial correlations, and comparable distributions of content categories for applicable subtests. Similar distributions of keyed responses, and stylistic features such as item formats, were also factors in item selection.

Each Form P subtest was constructed to have 50% of its content the same as that of the corresponding Form O subtest. This was accomplished by the use of common items drawn from Form O. Selection criteria were acceptable difficulty and discrimination statistics, balanced subject-matter content, balanced keyed response options, and an average item difficulty comparable to that of the full set of items in the Form O subtest. In addition, the positions of the common items were considered in the selection process in order to balance items taken from the beginning, middle, and end of each subtest.

The considerations for selecting common items also governed the selection of new items for Form P. The characteristics of Form O items to be replaced were particularly considered and item-by-item matches were attempted. A number of trade-offs were necessary to balance content, keys, appearance, and position; but in general, priority was given to matching the item difficulty and discrimination of Form P1 and Form P2 subtests.

Table 2 shows the experimental item data for the Form P selections. In the context of the analysis of experimental items (i.e., based on basic airmen and OTS cadet data), Forms P1 and P2 were highly similar in subtest difficulty, with no differences for 11 subtests and differences of .01 for five subtests. Item discrimination was exactly the same for P1 and P2 in four subtests, showed a difference of .01 for six subtests and differences from .02 to .04 for six subtests.

Table 2. Mean Item Difficulty (p) and Discrimination Statistics (r_{bis}) for Experimental Items Selected for Forms P1 and P2

Subtests	p P1	p P2	r_{bis} P1	r_{bis} P2
Verbal Analogies	.61	.60	.47	.47
Arithmetic Reasoning	.53	.53	.56	.56
Reading Comprehension	.64	.64	.54	.55
Data Interpretation	.52	.53	.45	.48
Word Knowledge	.56	.56	.55	.57
Math Knowledge	.58	.58	.55	.54
Mechanical Comprehension	.50	.49	.39	.39
Electrical Maze	.38	.38	.68	.67
Scale Reading (speeded)	.55	.55	.42	.38
Instrument Comprehension	.45	.44	.59	.62
Block Counting	.51	.51	.68	.65
Table Reading[a] (speeded)	-	-	.71	.70
Aviation Information	.43	.43	.59	.58
Rotated Blocks	.51	.51	.54	.54
General Science	.44	.44	.43	.40
Hidden Figures	.63	.63	.67	.68

[a]Item difficulty is not relevant for Table Reading. All items are inherently easy.

Comparison of the data for the new forms showed that high comparability, if not parallelism, was achieved between Forms O and P. The equivalence of the items in Forms P1 and P2 is discussed in the next section, which presents evidence regarding their psychometric properties when administered in the actual Form P context to applicants for officer commissioning training programs.

III. STATISTICAL ANALYSIS

Comparability of Forms P1 and P2

The data in this section were obtained from applicants for officer commissioning at operational test sites, including AFROTC detachments and Military Entrance Processing Stations, between 15 June 1987 and 31 October 1987. AFOQT Form P1 was administered to 3,216 applicants, and Form P2 to 2,976 applicants, using an equivalent groups design. Equivalent groups were assured by instructing test administrators to distribute Form P1 and then Form P2 to alternate examinees in each testing session. The last form distributed was recorded, and the next testing session was begun with the distribution of the other form. Scores were analyzed for central characteristics, item difficulty and discrimination, internal consistency reliability, and intercorrelation of the composites.

Score Characteristics

Common Items

Fifty percent of the items in each Form P subtest were drawn from Form O.[1] The score means and standard deviations of the items selected from Form O to appear in common in Forms P1 and P2 attest to the comparability of the two samples. Table 3 shows that the largest mean score difference for any subtest was less than one-third score point (Math Knowledge), and that the differences in standard deviations were similarly small. A similar pattern was found when subtest scores were combined into composite scores. The average performance of the samples tested on Form P1 and on Form P2 differed by about .5 point or less on four composites. The largest difference was 1.14 points on the Navigator-Technical composite.

Total Scores

Each AFOQT subtest is scored as one point for each correct answer. (There is no correction for guessing.) Table 4 shows the mean scores and standard deviations of the total test (new and common items) for Form P examinees. Form P2 has mean scores that are higher than those of Form P1 for nine subtests, with differences ranging from .01 point higher (Electrical Maze and Table Reading) to approximately 1.41 points higher (Reading Comprehension).

---

[1]After the Form P1 and P2 booklets were printed and distributed to operational test sites, two common items were determined to be inadequate in construction. These items -- one in the General Science subtest and one in the Aviation Information subtest -- remain in the test booklets but are omitted from the test scoring procedure.

Table 3. Score Means and Standard Deviations of Common Items on AFOQT Forms P1 and P2

Subtest	Number of common items scored	Form P1 Mean	Form P1 SD	Form P2 Mean	Form P2 SD
Verbal Analogies	13	8.18	2.44	8.35	2.44
Arithmetic Reasoning	12	7.34	2.81	7.10	2.83
Reading Comprehension	13	8.47	2.93	8.57	3.06
Data Interpretation	12	7.64	2.39	7.67	2.40
Word Knowledge	13	7.52	3.12	7.53	3.13
Math Knowledge	12	7.64	3.16	7.34	3.10
Mechanical Comprehension	11	5.62	2.36	5.60	2.39
Electrical Maze	10	3.99	2.11	3.92	2.03
Scale Reading	20	11.64	3.44	11.36	3.45
Instrument Comprehension	10	5.52	2.73	5.55	2.82
Block Counting	10	6.07	2.33	5.78	2.15
Table Reading	20	13.95	3.64	13.92	3.50
Aviation Information	9	4.22	2.24	4.28	2.22
Rotated Blocks	7	3.67	1.76	3.75	1.81
General Science	9	4.22	1.97	4.20	1.97
Hidden Figures	8	5.30	1.77	5.33	1.74
Composite					
Pilot	103	59.19	14.20	58.75	13.99
Navigator-Technical	131	77.10	18.63	75.96	18.26
Academic Aptitude	75	46.79	12.89	46.56	13.01
Verbal	39	24.17	7.26	24.46	7.44
Quantitative	36	22.62	7.09	22.11	7.11

Note. Form P1 N = 3,216; Form P2 N = 2,976.

The seven subtests in Form P1 that have higher mean scores than those of their counterparts in Form P2 show differences ranging from approximately .09 point higher for Word Knowledge to 1.37 points higher for Math Knowledge. Restated, these data indicate that Form P2 is slightly easier for nine subtests, particularly Reading Comprehension, and Form P1 is easier for seven subtests, particularly Math Knowledge. Because of the large sample sizes, even differences as small as .01 raw score point are statistically significant. The differences between means of Forms P1 and P2 for Reading Comprehension (1.41 score points), Math Knowledge (1.37 score points), and Block Counting (1.09 score points) may be meaningful.

When the subtests are organized into composites, Form P1 is shown to be slightly easier for the Pilot, Navigator-Technical, and Quantitative composites, with differences between means ranging from .61 to 3.34 score points. The Academic Aptitude and Verbal composites appear to be slightly easier in Form P2, with their means being .91 and 2.01 points higher, respectively, than those of Form P1. The largest difference between means was seen for Navigator-Technical (3.34 raw score points).

Table 4. Means and Standard Deviations of AFOQT Forms P1 and P2 Subtest and Composite Raw Scores

Subtest	Number of items scored	(Form P1 Mean	SD	Form P2 Mean	SD	t^a
Verbal Analogies	25	16.21	4.48	16.90	4.40	6.11
Arithmetic Reasoning	25	15.48	5.70	15.09	5.55	2.72
Reading Comprehension	25	15.17	4.99	16.58	5.24	10.82
Data Interpretation	25	16.00	4.50	16.67	4.83	5.64
Word Knowledge	25	15.16	5.74	15.07	5.78	0.61
Math Knowledge	25	16.63	6.24	15.26	6.09	8.81
Mechanical Comprehension	20	9.93	4.18	9.96	4.17	0.28
Electrical Maze	20	7.76	3.74	7.77	3.48	0.15
Scale Reading	40	23.79	6.86	22.51	6.59	1.64
Instrument Comprehension	20	11.52	5.27	11.98	5.45	3.37
Block Counting	20	12.79	4.34	11.70	3.99	10.30
Table Reading	40	27.59	7.11	27.60	6.84	0.06
Aviation Information	19	8.49	4.19	9.05	4.22	7.06
Rotated Blocks	15	8.00	3.21	8.47	3.40	5.58
General Science	19	9.34	3.94	9.15	3.67	1.96
Hidden Figures	15	9.99	3.12	9.78	2.95	2.72
Composite						
Pilot	204	118.08	28.35	117.47	27.74	0.86
Navigator-Technical	264	157.29	38.38	153.95	37.43	3.48
Academic Aptitude	150	94.65	25.43	95.56	26.05	3.07
Verbal	75	46.54	13.51	48.55	13.94	5.75
Quantitative	75	48.11	14.57	47.01	14.82	2.94

Note. Form P1 N = 3,216; Form P2 N = 2,976.
[a]Due to the large sample sizes, all differences between means for these samples are significant beyond the .01 level.

The standard deviation indicates the way measures are dispersed around a mean. One standard deviation on each side of the mean accounts for the scores of approximately two-thirds of any sample with normally distributed scores. For example, in interpreting Table 4, two-thirds of the Form P1 sample had Block Counting (BC) scores that differed from the mean score of 12.79 by 4.34 points (plus and minus); in other words, the said scores ranged from 8.45 to 17.13. Two-thirds of the P2 sample had BC scores ranging from 7.71 to 15.69. The scores achieved by the two samples in the range of the mean plus or minus one standard deviation are quite close, 8.68 points for Form P1 and 7.98 points for Form P2, indicating quite similar spreads around their respective mean scores.

Knowledge of the standard deviation of test scores is useful for a variety of applications. One of these is to compare the scores of samples on alternate test forms to see if one of the forms has a more restricted score range than the other. If this were the case, there could be different validity outcomes computed for the two forms. (Validity will be discussed in

a later section.) Forms P1 and P2 do appear to be comparable, however, with respect to the standard deviations of both the subtests and composites.

Skew and Kurtosis

Skew. When a distribution of scores is not symmetrically bell-shaped, it indicates that a larger proportion of the group scored either on the high end (negative skew) or the low end (positive skew) rather than in the middle. Appendix C indicates that, for the most part, Forms P1 and P2 show highly similar skewness. The differences between the P1 and P2 subtest coefficients of skewness range from zero (SR and AI) to .17 (VA). The differences between P1 and P2 composites range from .01 (N-T and AA) to .08 (V and Q). All of the composites are negatively skewed, indicating that higher scores were more frequent than lower ones.

Kurtosis. Kurtosis refers to the peakedness of a score distribution. Given the same number of cases, a tall and narrow curve is self-defined as a small range of scores; a short and wide curve, as a wide range of scores. Departures from normal kurtosis did not appear to be extreme. For the most part, coefficients of kurtosis are comparable for the P1 and P2 subtests, with the differences ranging from zero (AI) to .33 (TR). Differences in kurtosis of composite scores range from .01 (N-T) to .11 (Q).

Item Characteristics

Analyses were conducted on the data of the two samples referenced above to obtain item difficulty and item discrimination statistics. Appendix D shows the mean difficulty and discrimination values and the distributions of the item indices for each subtest in Form P1 (Tables D-1 through D-3) and in Form P2 (Tables D-4 through D-6).

Item Difficulty. The average item difficulty does not vary between each Form P1 and Form P2 subtest by more than .06 score point (Tables D-1 and D-4), and the difference is typically much smaller. Further, the distributions of item difficulty for the two forms are similar (Tables D-2 and D-5). These results compare very favorably to the data for the P1 and P2 items administered in the context of the experimental pool (see Section II above). The parallelism of the forms with respect to item difficulty appears to be closer in these data for the administration of Form P to officer applicants.

Item Discrimination. Test item analyses use the biserial correlation to measure the relationship between answering an item correctly and being in a top group of some type, usually a defined top sector of total scores. This correlation should be high for the keyed answer to a multiple-choice test question, and negative for all the wrong options. The samples of applicants for officer commissioning provided data indicating that the AFOQT subtests in Forms P1 and P2 had mean item-total test biserial correlations that more than met these criteria (Tables D-3 and D-6). Mean biserials (Tables D-1 and D-4) ranged from .51 to .72 in Form P1 (Scale Reading and Math Knowledge) and from .48 to .74 in P2 (Scale Reading and Instrument Comprehension). These data compare favorably with those collected for the same test items from basic airmen and OTS cadets in the experimental field tests. The differences between the mean biserials of the two forms were never greater than .06 for any subtest, providing further evidence of the parallelism of Forms P1 and P2.

Reliability of Subtests and Composites

Several testing concepts pertaining to the consistency of test measurements are associated with various types of reliability. One of these is stability of a measure administered more than once, estimated by correlating the scores obtained from a group of examinees administered the same test on two occasions. Another is "internal consistency reliability," which, in the case of the AFOQT, is computed by segmenting the subtests into items and computing their equivalence. A third is the "standard error of measurement" (SEM), a measure of the precision of a test score. The SEM indicates the area in which an examinee's "true" score is expected to fall, given obtained scores on repeated test administrations. An examinee's obtained score is expected to lie within one SEM on each side of the true score about two-thirds (68%) of the time. The smaller the SEM, the more confident one can be that different scores of two examinees on the same test represent true differences in aptitude. These different types of reliability in relation to the AFOQT Form P are discussed below.

Stability

The test-retest reliability is assumed for Form P from data for Form O, as the two forms are comparable. Arth (1986a) conducted a study of applicants for officer training, of whom 2,246 were retested on Form O between October 1981 and December 1983. The retest group differed in terms of the intervals between test and retest, with 312 retesting in less than 6 months, 1,300 in 6 to 11 months, 443 in 12 to 17 months, and 191 in 18 months or longer. For all retesters, the test-retest correlations on the composites were as follows: Pilot = .812, Navigator-Technical = .852, Academic Aptitude = .853, Verbal = .880, and Quantitative = .775. Higher or lower reliabilities might have been obtained had the estimates been computed separately for groups with different retest intervals. Allen and Yen (1979) noted that changes in lengths of time can affect test-retest reliabilities in different ways. Other factors which may have influenced the magnitude of the correlations are carry-over effects such as recall, practice, coaching, and motivation. Carry-over effects due to motivation are probable for the Arth (1986a) sample. The retesters were a self-selected group who requested an opportunity to retake Form O, presumably as an attempt to improve their scores to qualify for commissioning or aircrew training programs. Nevertheless, the reported reliabilities are moderately high and suggest stability in AFOQT measurements across time.

Internal Consistency

The Kuder-Richardson 20 (KR-20) formula was used to compute the reliability of the Form P1 and Form P2 subtests (Table 5), and a formula developed by Wherry and Gaylord (1943) was used to estimate the reliability of the composites (Table 6). As shown in Table 5, the reliabilities of the subtests range from .75 to .91 in Form P1 and from .71 to .90 in Form P2. The lower reliabilities were for Electrical Maze in Form P1 and for General Science in Form P2, and both high reliabilities were for Table Reading. (It should be noted that KR-20 reliabilities tend to be overestimated for speeded subtests and for the composites that contain them.) The comparability of the Forms P1 and P2 subtests with respect to reliability is striking. Six of the subtests--VA, AR, WK, MC, TR, and AI--have KR-20 reliabilities that differ by

only .010 or less between the two forms. The differences for RC, MK, EM, SR, IC, RB and HF range from .015 to .028. The largest differences are themselves inconsequential (.040, .044, and .048 for BC, DI, and GS, respectively). These findings indicate close comparability between Forms P1 and P2.

Table 5. Internal Consistency Reliability (KR-20) and Standard Error of Measurement (SEM)

Subtest	KR-20 P1	KR-20 P2	SEM P1	SEM P2
Verbal Analogies	.793	.791	2.025	2.010
Arithmetic Reasoning	.877	.867	2.001	2.024
Reading Comprehension	.826	.854	2.082	2.005
Data Interpretation	.772	.816	2.148	2.071
Word Knowledge	.876	.874	2.026	2.051
Math Knowledge	.902	.886	1.955	2.057
Mechanical Comprehension	.767	.770	2.020	2.000
Electrical Maze	.748	.720	1.878	1.839
Scale Reading	.851	.831	2.646	2.705
Instrument Comprehension	.876	.891	1.853	1.798
Block Counting	.835	.795	1.765	1.806
Table Reading	.908	.900	2.154	2.166
Aviation Information	.799	.806	1.878	1.859
Rotated Blocks	.751	.774	1.601	1.619
General Science	.762	.714	1.921	1.962
Hidden Figures	.766	.740	1.508	1.501

Table 6. Composite Reliabilities and Standard Errors of Measurement (SEM) Estimated From Subtest KR-20 and Communality Values

Composite	Reliability estimates				Standard error of measurement			
	KR-20		Communality		KR-20		Communality	
	P1	P2	P1	P2	P1	P2	P1	P2
Pilot	.958	.957	.867	.869	5.81	5.75	10.34	10.04
Nav-Tech	.971	.969	.914	.915	6.54	6.59	11.26	10.91
Academic Apt	.961	.963	.902	.913	5.02	5.01	7.96	7.68
Verbal	.931	.937	.834	.856	3.55	3.50	5.50	5.29
Quantitative	.941	.943	.845	.860	3.54	3.54	5.73	5.55

In the absence of a parallel forms reliability measure obtained from retesting the same examinees on both Forms P1 and P2, the Wherry-Gaylord (W-G) procedure was used to obtain an estimate of the reliability of each composite. The formula uses reliabilities for the subtests in a given

composite. Subtest reliabilities were computed using two methods. First, the KR-20 method was used to obtain an upper bound estimate of composite reliability. This estimate is upper bound because the KR-20 values are inflated due to the presence of both speed and mixed model subtests in the AFOQT, rather than power subtests exclusively. Second, communality estimates were computed. The communality of each subtest is its multiple correlation (R) with all other subtests. Communality was computed by regressing each subtest raw score against the raw scores for the remaining 15 subtests. The communality procedure provides a lower bound W-G estimate of composite reliability. The "true" reliability for each composite would be expected to fall between the estimates computed by the two methods.

The W-G estimates of the reliabilities for the Form P composites (see Table 6) are higher than those for their component subtests, ranging from .93 to .97 for Form P1 and from .94 to .97 for Form P2, based on the KR-20 method. Using the communality method, the estimates range from .83 to .91 in Form P1 and from .86 to .92 in Form P2. The Navigator-Technical composite shows the highest reliability in both forms, but all are acceptably high. The estimated composite reliabilities for the two forms are either identical (Pilot) or close to identical, with differences of .01 to .02.

Standard Error of Measurement (SEM)

The SEMs are highly similar for the Forms P1 and P2 subtests (Table 5). They range from approximately 1.5 to 2.2 score points except for Scale Reading, which has standard errors of 2.6 and 2.7 in P1 and P2, respectively. In practice, the SEM columns of Table 5 would be used to estimate the range of an examinee's "true score" on a particular subtest; that is, the obtained score minus and plus the SEM. Table 6 displays the SEMs for the composites. These are particularly important in that it is the composite scores which influence selection and placement. Form P1 SEMs tend to be higher than those of Form P2, but the differences are quite small, varying from .01 to .06 score point (KR-20 method) and from .18 to .35 score point (communality method). These findings add further support to the comparability of Forms P1 and P2.

Intercorrelations

Additional evidence is seen in the pattern of intercorrelations between composites for the two forms (Table 7). The correlations between any pair of composites for Form P1 did not differ from those for Form P2 by more than .01. Equivalent correlations were obtained for 7 of the 10 pairs. Results for subtest intercorrelations were somewhat more variable (Appendix E), although differences between the correlations for subtest pairs on Form P1 versus Form P2 rarely exceeded .05.

Standardization

Scoring Scales and Norms

To enhance the interpretability of results obtained by examinees, scoring scales and norms have been defined for the AFOQT. In the absence of scales and norms, raw test scores are not meaningful, in part because they

Table 7. Intercorrelations of Composite Raw Scores for AFOQT Forms P1 and P2

	Navigator-technical	Academic aptitude	Verbal	Quantitative
Pilot				
Form P1	.93	.75	.61	.74
Form P2	.93	.74	.61	.73
Navigator-Technical				
Form P1		.84	.62	.90
Form P2		.84	.62	.89
Academic Aptitude				
Form P1			.90	.91
Form P2			.90	.91
Verbal				
Form P1				.64
Form P2				.64

are affected by the difficulty of items on different test forms. On early forms of the AFOQT, raw scores were converted to a stanine scale (Valentine & Creager, 1961). However, the convention for the past three decades has been to derive percentile scales for the composites. Percentile scores on the AFOQT are expressed in terms of the percentage of examinees in the standardization sample who fall below a given raw score point.

The normative base for AFOQT Forms P1 and P2 is a sample of military personnel tested on Form N. As described by Gould (1978) in a report on the development and standardization of Form N, the 2,618-case sample consists of basic airmen (25.4%), AFROTC cadets (27.1%), AFA cadets (27.1%), OTS cadets (10.2%), and active duty second lieutenants (10.2%). The sample was designed to represent the full range of ability expected in the officer applicant population.

Equating

Form P scores were linked to the normative group using equipercentile equating to Form O scores. Form O had been equated to Form N in a previous study (Rogers, Roach, & Wegner, 1986). Analytic techniques were designed to achieve comparability to the Form N composite scales.

The Form P equating process proceeded in two steps. Initially, a set of provisional raw-to-percentile score conversions were developed on a research sample of random and equivalent groups of Forms O, P1, and P2 examinees enrolled in three military training programs: enlisted personnel in Basic Military Training and officer candidates in OTS and AFROTC (Steuck, Watson, & Skinner, 1988). The provisional conversions were used during the Initial Operational Test and Evaluation (IOT&E) of Form P conducted with officer

applicants between 15 June 1987 and 31 October 1987. Final conversion tables (see Appendix F) were produced, based on analyses of the applicants' data. These conversion tables were implemented in August 1988, when Form P officially replaced Form O as the operational AFOQT.

Validity

There are several approaches to the estimation of test validity. Predictive validity, the ability of a test score to predict future performance, is the pertinent measure for evaluating the AFOQT. The Pearson product-moment correlation is generally used to establish a test score's relationship to a performance measure. The higher the validity coefficient, the more accurate the prediction, given a representative sample of sufficient size to ensure reliable and stable results. Because score data are obtained for samples of a population, an obtained validity coefficient may be the result of chance factors such as sampling error which affect the composition of one particular sample, and therefore may not be a true indicator of the strength of the relationship between a test score and a performance measure. The probability (p) of obtaining a positive or negative correlation, when the actual correlation is zero, can be computed. If that probability is .05 or less, the result is considered to be statistically significant.

The predictive validity design requires a criterion of performance, whether in training or on the job, to relate to test scores. Various criterion groups are referenced in past AFOQT Form O validity studies. At this time, no research on the validity of Form P has been completed. The comparability demonstrated between Form P and Form O permits generalization of Form O's validity studies to Form P (Steuck et al., 1988).

Selection for Commissioning Programs

Validity research has shown the AFOQT Form O to be significantly predictive of training in the Officer Training School (OTS) and the Air Force Reserve Officer Training Corps (AFROTC) (Cowan, Barrett, & Wegner, 1989, 1990).

OTS. Passing scores on the AFOQT and a Bachelor's degree from an accredited university or college are required for application to OTS. Research was conducted to validate the OTS process of selecting college graduates for the 12-week training program (Cowan et al., 1990). The predictors studied included the percentile scores on the five AFOQT composites. Performance criteria were measures of success in OTS: (a) a dichotomy reflecting graduation or elimination from training, (b) final grade on academic material, (c) a training effectiveness rating from instructors, and (d) an indicator of whether or not the cadet was a distinguished graduate. Sample sizes varied by criterion (n = 3,200 to n = 4,500 cadets).

Most AFOQT composites were found to be significantly predictive of all criteria, with the highest correlations obtained between the Academic Aptitude and Verbal composites and final grade in training (r = .38 and r = .43, respectively). Validities for other criteria were lower, with most

ranging between .10 and .20 and significant at $p \leq .01$. The lower correlations were not unexpected due to the restricted variance of both the dichotomous criteria and the rating scale used by instructors. For example, the vast majority of cadets graduate from training and few qualify for distinguished graduate status. Another factor impacting the magnitude of the validities was restriction in range from use of the AFOQT to determine qualifications for the officer force. OTS cadets are a highly select subgroup of applicants for commissioning training. Reported correlations were not corrected for range curtailment on the test. Overall, the finding that test scores consistently related to training performance attest to the validity of the AFOQT as an OTS selector.

AFROTC. The AFROTC administers a 2-year Professional Officer Course (POC) to prepare accepted candidates for officer duty. POC candidates are evaluated by a Quality Index Score (QIS), which consists of six differentially weighted variables. These are the Verbal, Quantitative, and Academic Aptitude composites of the AFOQT; the Scholastic Aptitude Test (SAT) score; cumulative grade point average (GPA); and a Detachment Commander's rating, which is an overall rating based on the first five variables and a personal interview. Research was conducted to validate these predictors against training and post-commissioning performance variables (Cowan et al., 1989).

Validity analyses for the AFOQT composites were reported for samples of 1,000 to 9,500 AFROTC cadets. The criteria were completion/non-completion of the POC; an instructor rating of training performance; distinguished graduate status; and three experimental supervisory ratings of job performance, potential for career progression, and job motivation.

Study results showed that every criterion except job motivation was significantly predicted by one or more of the AFOQT composites. The Academic Aptitude composite offered the greatest predictability of the three AFOQT composites, with correlations near .15 for instructor rating, distinguished graduate, and potential for career progression criteria. The correlations obtained for AFROTC, like those observed for OTS performance, are likely to be substantial underestimates of the predictive power of the AFOQT. Approximately 25% of the QIS used for AFROTC cadet selection decisions was due directly to AFOQT scores. In combination, selection on AFOQT aptitudes and similar ones measured by the SAT produces a fairly homogeneous group of cadets. The impact of aptitude curtailment was not accounted for in the AFROTC validities.

Rated and Non-Rated Training

The criterion-related validity of the AFOQT as a selection instrument for Undergraduate Navigator Training (UNT), Undergraduate Pilot Training (UPT), and non-aircrew technical training has been established by several studies, summarized below.

Undergraduate Navigator Training (UNT). The Navigator-Technical composite is one of the primary selectors for UNT. A study by Shanahan and Kantor (1986) evaluated all the AFOQT composites against four UNT standards. The criterion measures were training outcome (graduation/elimination),

average classroom lesson score, average simulator lesson score, and average flying lesson score. All of the AFOQT composites were found to correlate significantly with the graduation/elimination performance measure. The correlations for the composites with that measure, uncorrected for range restriction on the tests, were between .10 and .19, all significant at $p \leq .01$.

UNT classroom lesson and UNT simulator lesson grades were also well predicted by the composites, especially by Quantitative (uncorrected $r = .40$ and .32, respectively). The Pilot and Navigator-Technical composite scores were the most predictive for UNT flying lesson grades (uncorrected $r = .21$ and .24, respectively).

The results of hierarchical regression analyses provided evidence that when Navigator-Technical scores were combined with Quantitative scores, the prediction of the various performance criteria was improved. The multiple correlations (R) thus obtained were .20 for graduation/elimination, .43 for classroom lesson grades, .34 for simulator lesson grades, and .26 for flying lesson grades.

A study of the validity of AFOQT subtests and composites for 632 UNT attendees was conducted by Arth, Steuck, Sorrentino, and Burke (in preparation). The AFOQT subtests with the highest correlations (corrected for restriction of range) with a UNT pass/fail measure were MK (.21), AR (.23), BC (.24), and SR (.25). The Pilot, Navigator-Technical, and Quantitative composites had corrected correlations with UNT outcome of .25, .27, and .23, respectively.

Undergraduate Pilot Training (UPT). Candidates for OTS and UPT who do not have a private pilot license must successfully complete the Flight Screening Program (FSP) prior to entering OTS. The validity of the Pilot composite in predicting UPT outcome for 719 attendees of FSP and UPT between 1982 and 1986 was investigated by Sawin (1987). The AFOQT Form O Pilot composite percentile score had a statistically significant correlation with UPT outcome (uncorrected $r = .15$, $p < .001$).

Arth et al. (in preparation) studied the validity of the various AFOQT subtests and composites in predicting UPT outcome for 695 officers. The subtests that best predicted the UPT pass/fail outcome were Instrument Comprehension (corrected $r = .38$) and Aviation Information (corrected $r = .30$), both of which also have face validity for pilot tasks. The Pilot and Navigator-Technical scores were the best predictors among the composites of UPT outcome, with corrected r's of .30 and .24, respectively.

Non-Rated Technical Training. AFOQT data were obtained for 9,029 non-aircrew officers who attended 29 entry-level and 8 advanced non-rated technical training courses (TTC) between October 1979 and December 1983 (Arth, 1986b). The AFOQT composites, categorized as either Rated (Pilot and Navigator-Technical) or Non-Rated (Academic Aptitude, Verbal, and Quantitative) were all correlated with final TTC grades, although only the non-rated composites are used for selection in this context. Most of the course grades were well predicted by all composites, entry-level courses particularly. The correlations between composites and grades ranged from

.01 to .62, with the most frequent intervals being .31 to .40 (59 correlations) and .41 to .60 (57 correlations). The Academic Aptitude composite, with 28 correlations of .31 to .53, predicted course grades best, on the whole. The Navigator-Technical composite was the next best predictor, with 24 correlations that ranged from .31 to .62. The correlations reported were uncorrected for the range restriction in test score variance resulting from preselection of the officer samples on AFOQT measures.

Improvements in prediction were achieved for the non-rated composites by combining them. For 20 of the 37 courses, a linear combination of Verbal and Quantitative scores predicted final grades significantly better than did the single composites. Another noteworthy result was that in the cases of Air Traffic Control and one Intelligence class, the Pilot and Navigator-Technical scores predicted grades better than did the non-rated composites, singly or in combination.

Air Weapons Controllers. The development of a selection strategy to improve the success rate of Air Weapons Controllers during training and in field assignments entailed obtaining predictive validity statistics for 968 students at five organizations responsible for training Air Weapons Controllers (Finegold & Rogers, 1985). Pearson product-moment correlations were computed between the AFOQT composites and criteria consisting of academic grade, success (i.e., pass or fail), and student class rank. All of the AFOQT composites were significantly related to all of the performance criteria at the $p < .01$ level. Academic Aptitude was the most consistently successful composite in predicting student performance, with its correlations ranging from .26 to .38. The Navigator-Technical composite was the next most successful predictor, with correlations from .23 to .39. The performance criterion best predicted by all composites was class ranking.

Validity Generalization for Non-Rated Officers. As a preliminary attempt to assess the comparability of AFOQT validity across test forms, Hartke and Short (1988) conducted a Schmidt-Hunter meta-analysis on 47 different validity coefficients obtained for AFOQT Academic Aptitude composite scores against technical training grades. The Schmidt-Hunter procedure calculates a weighted mean validity coefficient corrected for study differences in range restriction and predictor/criterion reliability, and variance due to sampling error, among other differential effects. The weighted mean validity coefficient is assumed to be the best estimate of true validity if the variance in correlation coefficients attributable to sampling error and other study effects accounts for most ($\geq 75\%$) of the observed variance in the set of effects.

The AA composite score data were collected from studies in which either several AFOQT forms were involved (L, M, N, and O) or there was no information as to which form was used. Therefore, in this preliminary study, the question of validity generalizability across forms could not be answered. Predictor and criterion reliability data and score variance were also not available, precluding corrections for these factors. Nevertheless, the meta-analysis could be conducted to determine the generalizability of AA across all jobs in the sample, and subgroups of jobs for major occupational groupings.

The results of this study indicated that true validity was not the same across all job titles, because although a weighted mean correlation of .39 was obtained, sampling error accounted for only 33% of the variance. The wide range of job occupations in the sample probably accounted for this result to a large extent. For the four major occupational subgroups studied, observable sampling error variance was higher, but exceeded 75% for only one subgroup studied, Intelligence and Security Police. The meta-analysis for this subgroup produced a weighted mean correlation of .44, with 100 percent of the observed variance attributable to sampling error. At this time, then, the AA composite's validity can be considered to be the same for all occupations within only the Intelligence and Security Police subgroup. The results of this study also indicate that although AA validity varies across Air Force specialties, the individual true validities will vary around the mean weighted validity coefficient of .39. Planned meta-analysis research on the AFOQT will include all the composites, larger data sets, more complete information on test forms and test statistics, and more homogeneous job title groups.

Construct Validity

In psychological testing, a "construct" is a meaningful attribute (e.g., anxiety; verbal ability) that is hypothesized to operate dimensionally in human behavior. Thus, there are supposedly differing levels of "anxiety" and "verbal ability." Construct validity can support the existence of such theoretical dimensions through several techniques that demonstrate how tests hypothesized to measure the same construct will group together. Whereas the predictive validity of a personnel selection test such as the AFOQT provides the most compelling evidence to support inferences regarding the test's utility, construct validity can confirm the nature of the attribute being measured. A variety of statistical evidence supports the constructs of the AFOQT subtests and composites. These include factorial validity of the subtests, intercorrelations of relatively homogeneous subtests, and correlations with non-military tests and measures.

Factorial Validity. A factor analysis of AFOQT Form O was conducted for a random sample of 3,000 examinees to determine the dimensions underlying the 16 subtests (Skinner & Ree, 1987). Five factors identified in the results strongly suggested the factor names of Verbal, Quantitative, Space Perception, Aircrew Interest/Aptitude, and Perceptual Speed.

Verbal Analogies, Reading Comprehension, and Word Knowledge were the tests that clearly defined the Verbal Factor (i.e., had the highest loadings on Factor I). Math Knowledge, Arithmetic Reasoning, and Data Interpretation clearly identified the Quantitative factor. Factor III, Space Perception, was composed of Rotated Blocks, Block Counting, Electrical Maze, Hidden Figures, and Mechanical Comprehension. Factor IV, Aircrew Interest/Aptitude was best defined by Aviation Information and Instrument Comprehension. The Perceptual Speed factor included the speeded subtests, Table Reading and Scale Reading. The AFOQT test constructs identified by these factor names appear to have been well validated by the types of tests that formed each factor.

Subtest Intercorrelations. Subtest intercorrelations for AFOQT Form P, obtained from officer applicant samples of 3,216 for Form P1 and 2,976 for Form P2, varied from .16 to .76. With samples this large, even the smaller correlations are significant and indicate that all the subtests are measuring something in common. The commonalities might be due to general ability and the basic verbal ability required to read test instructions, together with the ability to follow those instructions in detail. What is of interest with respect to construct validity is whether tests measuring similar attributes have high correlations with each other, but low correlations with measures of hypothetically dissimilar aptitudes. The subtest intercorrelation matrices for Forms P1 and P2 in Appendix E provide evidence of such validity for the constructs suggested by the AFOQT factor analysis.

The verbal aptitude construct in the AFOQT appears directly or by implication in the subtest names of Verbal Analogies (VA), Word Knowledge (WK), and Reading Comprehension (RC). The highest intercorrelations obtained for these three subtests are with each other. The lowest correlations for these verbal subtests are with Electrical Maze (EM), a subtest that one can intuitively perceive as requiring little verbal power for its accurate responses. The verbal construct seems to be well established. Relatively high correlations are often observed between verbal subtests and math or scientific types of subtests on multiple aptitude batteries. It can be convincingly argued that math and science problems do require verbalization to some degree; hence, the moderately high correlations obtained between these subtest types on Form P.

The quantitative aptitude construct suggested by the factor analysis was defined by the Arithmetic Reasoning (AR), Data Interpretation (DI) and Math Knowledge (MK) subtests. These subtests tend to correlate more highly with each other than with other subtests. The lowest correlations of AR, DI, and MK are with Aviation Information (AI) (.20 to .34). The AI test is a knowledge test that does not require mathematical reasoning.

The construct identified as aptitude for space perception relies on five subtests for its definition: Electrical Maze (EM), Block Counting (BC), Rotated Blocks (RB), Hidden Figures (HF), and Mechanical Comprehension (MC). Appendix E indicates that these subtests in both Forms P1 and P2 correlate highly with each other ($< .39$), and with the Scale Reading subtest, which also measures visual-perceptual skills. The lowest correlations typically found for the tests defining the space perception construct were with WK and AI, subtests that clearly do not require spatial-perceptual skill.

The construct of aircrew interest/aptitude is supported by the intercorrelations of MC, IC, AI, and GS. These range from .45 to .65. The content of these subtests appears to be appropriate to the interests and aptitudes involved in successful aircrew work. Certain subtests have low correlations with AI but have moderate to high correlations with the other subtests supporting this construct (HF, SR, MK, and EM). The aircrew interest/aptitude construct appears to be a complex one that does not exclude to a great extent the other aptitudes measured by the AFOQT.

Perceptual speed was the last construct suggested by the factor analysis. The subtests whose descriptions support this construct, Table Reading (TR) and Scale Reading (SR), were well correlated (r = .53 and .55 on Form P1 and Form P2, respectively); but SR, which has computational components, was more highly correlated with subtests involved in the quantitative construct. TR also correlated well with BC, probably in part due to BC's perceptual aspect. Also, the BC test has been shown to have a speeded component (Skinner & Ree, 1987). The subtests with the lowest correlations with TR and SR were AI and WK, neither of which requires perceptual speed to respond to its items.

For the most part, data from the subtest intercorrelations show validity for the AFOQT constructs in that subtests requiring similar skills tended to be more highly correlated with each other than with other subtests, and in that the lowest intercorrelations tended to be between subtests that purport to require few, if any, of the same skills.

Correlations with Civilian Tests. A study was conducted to determine the extent to which the AFOQT composites correlate with similar measures in the Scholastic Aptitude Test (SAT) and the ACT composite of the American College Testing Program (Diehl, 1986). The subjects were AFROTC cadets entering the Professional Officer Course during Fiscal Year 1985 (FY85). Analyses were conducted for 1,907 cadets tested on both the AFOQT and SAT and for 1,240 cadets tested on both the AFOQT and ACT.

The highest correlations for the AFOQT composites were found between the Academic Aptitude composite and similar composites derived from the cadets' combined performance on the verbal and math sections of the civilian batteries (Table 8). The AFOQT Academic Aptitude composite showed a correlation of .80 with the ACT Composite and with the SAT Composite. Correlations between the part scores were only slightly lower. The AFOQT Verbal and Quantitative composites correlated .77 and .71, respectively, with their SAT counterparts SAT Verbal and SAT Math. Scores on the Navigator-Technical composite, which includes the subtests of the Quantitative composite, also related highly to performance on the SAT Math composite (r = .61). Some correlations were substantially lower, as might be expected. For example, the Pilot and Navigator-Technical composites did not correlate highly with the SAT and ACT Verbal composites, probably because there is little overlap in test content.

Table 8. Correlations of AFOQT Form O Composites with Civilian Tests

AFOQT composite	ACT composite	SAT composite	SAT verbal	SAT math
Pilot	.416	.400	.237	.449
Navigator-Technical	.566	.547	.326	.614
Academic Aptitude	.804	.801	.716	.687
Verbal	.681	.702	.772	.473
Quantitative	.683	.660	.429	.710

The high correlations between the AFOQT composites and their content-parallel composites in prominent civilian tests provide additional confirmation of the construct validity of the AFOQT.

IV. ADMINISTRATION AND SCORING

Administration

Testing Locations

AFOQT testing locations include Military Entrance Processing Stations (MEPS), Air Force Reserve Officer Training Corps detachments on university campuses, and Consolidated Base Personnel Offices (CBPOs) on military installations. The annual testing load is approximately 35,000 to 40,000 examinees.

Administrators

Test administration is conducted and monitored by Personnel Technicians (Air Force Specialty Code 73270). Qualifications for the Test Control Officers (TCOs), who supervise and monitor on-site testing programs, include attainment of the rank of Master Sergeant or higher. Test Examiners (TEs) responsible for actual administration must be at least Staff Sergeant selectees. Training activities for TEs include familiarization with testing regulations (AFR 35-8, Air Force Military Personnel Testing System, and AFR 30-17, Safeguarding CONTROLLED ITEM (Test Material) Information) and review of the manual for administration (Department of the Air Force, 1987b). The manual specifies requirements for the testing environment and preparation of testing materials. The local TCO supervises practice testing sessions with the TE. TCOs at Air Force major commands, as well as the Inspector General, ensure local TCOs comply with standards for testing conditions and administration procedures as specified in AFR 35-8 and the manual for administration.

Standardized Procedures

Information Pamphlet. Prior to testing, examinees receive an Information Pamphlet (Department of the Air Force, 1987a) to acquaint them with the nature of the AFOQT. The pamphlet describes the purpose, content, and schedule of the test and presents five sample items with answers for each subtest.

Test Environment. A space of 15 square feet (inclusive of aisles) per examinee is required. Desks are arranged for efficient test distribution and collection. Minimum standards of lighting and ventilation must be met, and all examinees must be visible to the TE. If waivers are requested for testing facilities that are below standard, the request must include documentation which describes the actions to be taken to counteract the deficiences. If test interruptions such as power failure occur, the TE provides instructions to the examinees with respect to handling of test booklets and test rescheduling. In such cases, the TE collects and safeguards the test materials.

Directions. Examiners at all testing sites are directed to deliver the test directions exactly as stated in the manual for administration. Each testing session begins with the TE's verifying each examinee's identity and having the examinee sign a test roster. The TE then distributes all necessary testing materials (i.e., test booklet, answer sheet, and pencil). The test purpose and rules and schedule of testing are explained. Each examinee is requested to certify, by signing the answer sheet, that he or she is physically fit to take the examination. Otherwise, the examinee is excused and rescheduled for a later testing date. Following collection of biographic and military personnel data, a Privacy Act Statement is read. This statement describes the authority to conduct AFOQT testing (10 United States Code 508, 509, and 510, and Executive Order 9397), the principal purpose for collection of personal identity information including Social Security Number, and routine uses of the scores. Examinees are informed that disclosure of personal information is voluntary but that refusal could result in denial of a military commission. Finally, testing commences with subtests being administered according to the schedule shown in Table 9. The total amount of time required to administer the AFOQT is approximately 4.5 hours.

Table 9. Testing Schedule

	Administration time (in minutes)	Testing time (in minutes)	Total time (in minutes)
Pretest Activities	24		24
Verbal Analogies	1	8	9
Arithmetic Reasoning	1	29	30
Reading Comprehension	1	18	19
Data Interpretation	1	24	25
Word Knowledge	1	5	6
Math Knowledge	1	22	23
Break	10		10
Mechanical Comprehension	1	22	23
Electrical Maze	3	10	13
Scale Reading	3	15	18
Instrument Comprehension	3	6	9
Block Counting	2	3	5
Table Reading	2	7	9
Aviation Information	1	8	9
Rotated Blocks	2	13	15
General Science	1	10	11
Hidden Figures	2	8	10
Collection of Materials	2		2
TOTAL TIME REQUIRED	1 hr 2 min	3 hrs 28 min	4 hrs 30 min

Test Security

The Office of Primary Responsibility (OPR) for test administration, scoring, and reporting of results for the AFOQT is the Personnel Testing Branch, Personnel Measurement Division, Headquarters Air Force Military Personnel Center (HQ AFMPC/DPMYOT), Randolph AFB, Texas. Administrative direction and oversight are provided to about 500 TCOs assigned throughout the Continental United States and overseas.

Test material and data are treated as confidential and are released only to authorized personnel. Further, provisions are made to prevent test compromise and to ensure scores are not obtained by fraudulent means. TCOs and TEs are fully advised of their responsibility for securing materials and scores and of potential punitive action that may ensue due to test compromise or loss of materials. All test administration personnel are informed that they are subject to administrative/disciplinary action under the Uniform Code of Military Justice (UCMJ) as specified in AFR 30-17, which relates to test compromise or loss of materials. TCOs and TEs may not take a test which they have administered, or to which they have access, without prior approval from HQ AFMPC/DPMYOT or until 6 months have elapsed since access to the test.

TCOs and TEs are responsible for safeguarding and/or retaining personal control of test materials at all times and must adhere to regulatory procedures for reporting any examinee misconduct detected. Examinee activities that constitute test compromise are described to examinees prior to testing.

Scoring

AFOQT Scoring Procedure

Official scores used for Air Force personnel decisions are derived and reported by HQ AFMPC/DPMYOT to recruiters, detachment commanders, and other authorized users. Examinees record their test responses on machine-scannable answer sheets. After testing, TEs forward the answer sheets from the testing site for processing at a central facility maintained by HQ AFMPC/DPMYOT. Scoring accuracy is verified by periodic checks of the calibration of the scanning device, and of associated scoring and reporting software, using two procedures: (a) processing of "dummy" answer sheets with known scores, and (b) hand-scoring of examinees' answer sheets. Due to the proven accuracy of the computerized processing procedures, requests for hand-scoring and rescoring are usually denied. In anticipation of the possibility that scores might be challenged, however, answer sheets are retained for a period of 6 months after testing, along with microfiche records and tape files of scores and item responses.

The turnaround time for official scoring and reporting is about 2 weeks. In the interim, recruiters at the MEPS are authorized to use an alternative scoring procedure to estimate how well an applicant will perform on the AFOQT. The scoring procedure approved for use with Form P is called "Quick Score."

Quick Score

Prescreening of applicants for OTS or the AFROTC is an undertaking to reduce the costs associated with applicant processing. Recruiters can use the unofficial "Quick Score" estimates to make decisions about how to manage applicants while waiting to receive official AFOQT results from HQ AFMPC/DPMYOT. For example, the applications of high-scoring examinees who are most likely to meet Air Force aptitude entry requirements may be expedited.

Several early versions of a prescreening test were previously developed for predicting AFOQT scores. The Air Force Precommissioning Screening Test (AFPST), a screening test for selection to navigator training and to the Air Force Academy Preparatory School, was a short version of the Officer Quality composite (now known as the Academic Aptitude composite of the AFOQT) (Valentine, 1961). The AFPST, renamed the Pre-Enrollment Test, was revised and implemented in 1965, and underwent further revision in 1967 (Miller, 1966, 1968). The Pre-Enrollment Test was discontinued in April 1969.

The need for a prescreening device recurred in the early 1980s with the implementation of AFOQT Form O. Unlike the AFPST and the Pre-Enrollment Test, which were separate and shorter versions of the AFOQT, the Officer Screening Composites developed for Form O used a subset of the items in the operational battery (Rogers, 1985).

The Officer Screening Composites were used from July 1982 until August 1988, when Form P became the operational officer selection test. At that time, HQ AFMPC/DPMYOT requested that the applicant prescreening procedure be updated. Research was initiated by the Air Force Human Resources Laboratory to improve the prescreening procedure for Form P.

The development of the Form P Quick Score is described at length by Sperl (1988) and by Sperl and Ree (1989). Briefly, two techniques for selecting a subset of items from the AFOQT battery for the Quick Score composites were compared for predictability of scores on the full-length test and for gender and ethnicity effects. One technique identified items from each subtest by random selection; the second identified items with the highest item-subtest point biserial correlations, a strategy similar to that used to construct the Form O Officer Screening Composites.

Results favored the random item selection methodology. Although the two techniques (random and point biserial) both yield Quick Score composites which correlate well with examinees' scores on the corresponding composite derived from the full-length test, the random item selection technique shows considerably less bias in prediction for gender and ethnic groups. Another improvement offered by the use of the Form P Quick Score composites over the Form O Officer Screening Composites concerns the representativeness of test coverage. Selected subtests were excluded in the development of the Form O prescreening device; the Quick Score for Form P represents all subtests in proportion to their relative contributions to the total item count on the full-length test.

Quick Score composites based on the random item selection technique were recommended for implementation to HQ AFMPC/DPMYOT. Additional analyses were conducted to express or convert Quick Score composites to the percentile metric used to report results on the five full-length AFOQT composites. The needs of recruiters, who are the primary users of the Quick Score composites, were a major consideration. Recruiters had expressed dissatisfaction with the Form O conversion method, which used a confidence interval method to establish the range of percentile scores within which raw scores from short-form scoring were expected to fall (Rogers, 1985). The alternate procedure used for Form P is an expectancy table. It reports the percentage of examinees obtaining a given value on the Quick Score composite who are expected to reach or exceed selected percentiles on the corresponding AFOQT composite. Manuals were also developed to instruct TEs on the correct procedures for computing the Quick Score composite scores and to guide recruiters in the proper interpretation of the scores using the expectancy tables (Department of the Air Force, 1989a, 1989b).

Qualifying Scores. In the first stage of applying to AFROTC or OTS, all examinees take all of the AFOQT. The minimum AFOQT composite scores needed to qualify for acceptance are the 15th percentile for Verbal (V) and the 10th percentile for Quantitative (Q). To qualify for pilot/helicopter training, an applicant must achieve a minimum Pilot Composite (P) score at the 25th percentile and a minimum Navigator-Technical Composite (N-T) score at the 10th percentile. Also, the examinee's combined P and N-T scores must be at least at the 50th percentile (AFR 51-4). That is, if only the minimum score is achieved on one of the two composites, the applicant must compensate with a higher score on the other composite in order to achieve the 50th percentile on the combined scores.

To qualify for navigator training, the applicant must achieve a minimum N-T score corresponding to the 25th percentile, a minimum P score at the 10th percentile, and a combined N-T and P score at the 50th percentile.

Reporting Scores. Scores are reported to the test-taker verbally or in the form of a letter, or by a copy of a Report on Individual Personnel (RIP). Individual CBPOs report scores to Active Duty Air Force, National Guard, and Air Force Reserve personnel. The recruiter at a MEPS reports scores to the civilian applicants. The scores of ROTC cadets are sent to HQ AFROTC.

Score Interpretation. The selection and commissioning of officers from OTS and AFROTC are directly affected by the operational use of AFOQT scores. Proper understanding and utilization of these scores is therefore of high importance. The predictive information provided by AFOQT scores is combined with other data (such as physical fitness and educational attainment, among others) that help to determine personnel decisions. The AFOQT provides information regarding a candidate's aptitudes for a specified program. In some programs, failing to achieve an established AFOQT minimum score can be a sufficient reason for rejecting a candidate (Arth, 1985).

Extensive studies have determined the specific combinations of subtests that maximize the prediction for certain types of training. Computer scoring yields the five AFOQT composites that represent these subtest

combinations. With respect to interpretation, Air Force personnel who use the scores to make selection decisions about officer applicants receive a briefing that includes composite descriptions, the meaning of percentiles, and the score levels that are considered competitive. Test-takers who request score interpretation receive an explanation of their percentiles; that is, where they rank relative to all other examinees in the normative sample on the various composites.

Because of the way in which the AFOQT is constructed, it is possible to compare the score differences of different examinees on the same composite, and those of the same examinee on different composites. Care is taken in inferring strengths and weaknesses from the score differences, however. Because the score differences may be a result of chance, their proportion in excess of chance must be estimated before diagnostic inferences are drawn.

V. SUPPLEMENTAL SOURCES OF AFOQT INFORMATION

Points of Contact for the AFOQT

The three major Air Force agencies involved on a day-to-day basis with the AFOQT testing program are listed below. The functional responsibilities of the agencies differ as shown.

1. HQ United States Air Force
   Directorate of Personnel Plans
   Policy Division
   Washington, DC 20330

Develops plans, policies, and procedures to implement and monitor the AFOQT. Establishes and coordinates testing requirements and procedures needed to implement Air Force officer procurement policies.

2. HQ Air Force Military Personnel Center (AFMPC)
   Personnel Measurement Division
   Randolph AFB, TX 78150-5000

Executes test policy and oversees the operational program (test printing, distribution, administration, scoring, reporting, security).

3. Air Force Human Resources Laboratory (AFHRL)
   Manpower and Personnel Division
   Brooks AFB, TX 78235-5601

Performs research to support the operational test and testing program. Is responsible for test development and revision. Develops, analyzes, and evaluates tests and associated materials (e.g., Administration Manual, Quick Score). Conducts continuing research and development to ensure the use of the AFOQT and its psychometric adequacy for officer selection and classification decisions.

AFOQT Bibliography

In addition to the references cited in this manual, a bibliography of selected references relative to the AFOQT is given in Appendix G. Most of the references were extracted from a longer list of publications compiled by Cowan and Sperl (1989). Their paper covers 50 years (1937-1986) of research on selection and classification of U.S. military officers, encompassing many topics in addition to the AFOQT.

The AFOQT bibliography is arranged chronologically by report date from 1950 (about the time that Form A was introduced) to the present, providing material of interest both to those readers interested in gaining an historical perspective on the test and to those concerned primarily with the recent or current officer testing program. Most of the reports focus on AFOQT development, standardization, and validation topics.

Readers may obtain copies of the reports from:

(Qualified requesters--Department of Defense personnel and registered DTIC users)

Defense Technical Information Center (DTIC)
Cameron Station
Alexandria, VA 22304-6145
Commercial (202) 274-7633
AUTOVON 284-7633

(Private sector requesters)

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Commercial (703) 487-4650

References

Air Force Regulation (AFR) 30-17. (23 Jun 89). Safeguarding Controlled Item (Test Material) Information. Washington, DC: Department of the Air Force.

Air Force Regulation (AFR) 35-8. (25 Apr 83). Air Force military personnel testing system. Washington, DC: Department of the Air Force.

Air Force Regulation (AFR) 51-4. (25 Aug 86). Application procedures for undergraduate flying training (UFT). Washington, DC: Department of the Air Force.

Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company.

Arth, T.O. (1985). United States Air Force Officer Qualifying Test Manual for interpretation (updated through AFOQT Form O). Unpublished manuscript.

Arth, T.O. (1986a). Air Force Officer Qualifying Test (AFOQT) Retesting effects (AFHRL-TP-86-8, AD-A168 926). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Arth, T.O. (1986b). Validation of the AFOQT for non-rated officers (AFHRL-TP-85-50, AD-A164 134). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Arth, T.O., Steuck, K.W., Sorrentino, C.T., & Burke, E.F. (in preparation). Air Force Officer Qualifying Test (AFOQT): Predictors of undergraduate pilot training and undergraduate navigator training success. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Berger, F.R., Gupta, W.B., Berger, R.M., & Skinner, J. (1988). Air Force Officer Qualifying Test (AFOQT) Form P: Test construction (AFHRL-TR-88-30, AD-A200 678). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Brown, D.C. (1987). Military officers: Commissioning sources and selection criteria (HumRRO FD-PRD-87-42). Alexandria, VA: Human Resources Research Organization.

Cowan, D.K., Barrett, L.E., & Wegner, T.G. (1990). Air Force Officer Training School selection system validation (AFHRL-TR-89-65). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cowan, D.K., Barrett, L.E., & Wegner, T.G. (1989). Air Force Reserve Officer Training Corps selection system validation (AFHRL-TR-88-54). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cowan, D.K., & Sperl, T. (1989). Selection and classification of United States military officers: A fifty-year bibliography (1937-1986) (AFHRL-TP-88-45, AD-A205 631). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Department of the Air Force (1987a, January). Air Force Officer Qualifying Test information pamphlet (AFPT 997). Randolph AFB, TX: Air Force Military Personnel Center.

Department of the Air Force (1987b, January). United States Air Force Officer Qualifying Test: Manual for administration, Form P, versions 1 & 2 (AFPT 983). Randolph AFB, TX: Air Force Military Personnel Center.

Department of the Air Force (1989a, January). United States Air Force Officer Qualifying Test: Procedures for quick scoring Forms P (A guide for test control officers) (AFPT 988). Randolph AFB, TX: Air Force Military Personnel Center.

Department of the Air Force (1989b, January). United States Air Force Officer Qualifying Test: Interpreting quick score composites for Forms P (A guide for recruiters) (AFPT 989). Randolph AFB, TX: Air Force Military Personnel Center.

Diehl, G.E. (1986). Correlations among SAT, ACT, AFOQT, and grade point average. Unpublished manuscript.

Finegold, L.S., & Rogers, D.L. (1985). Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training (AFHRL-TR-85-13, AD-A158 162). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Gould, R.B. (1978). Air Force Officer Qualifying Test Form N: Development and standardization (AFHRL-TR-78-43, AD-A059 746). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hartke, D.D., & Short, L.O. (1988). Validity of the academic aptitude composite of the Air Force Officer Qualifying Test (AFOQT) (AFHRL-TP-87-61, AD-A194 753). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Miller, R.E. (1966). Development of officer selection and classification tests - 1966 (PRL-TR-66-5, AD-639 237). Lackland AFB, TX: Aerospace Medical Division, Personnel Research Laboratory.

Miller, R.E. (1968). Development of officer selection and classification tests - 1968 (AFHRL-TR-68-104, AD-679 989). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Miller, R.E. (1969a). Interpretation and utilization of scores on the Air Force Officer Qualifying Test (AFHRL-TR-69-103, AD-691 001). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Miller, R.E. (1969b). Interpretation and utilization of scores on the Air Force Officer Qualifying Test (AFHRL-TR-69-103, AD-691 001). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Miller, R.E. (1974). Development and standardization of the Air Force Officer Qualifying Test Form M (AFHRL-TR-74-16, AD-778 837). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Roach, B.W. (1986). The influence of college education on standardized test performance: Should multiple conversion scales be used? Paper presented at the 10th Symposium of Psychology in the DoD, US Air Force Academy, Colorado Springs, CO.

Roach, B.W., & Rogers, D.L. (1982). Development of the common metric. Paper presented at the 8th Symposium of Psychology in the DoD, US Air Force Academy, Colorado Springs, CO.

Rogers, D.L. (1985). Screening composites for Air Force officers (AFHRL-TP-85-2, AD-A154 315). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Rogers, D.L., Roach, B.W., & Short, L.O. (1986). Mental ability testing in the selection of Air Force officers: A brief historical overview (AFHRL-TP-86-23, AD-173 484). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Rogers, D.L., Roach, B.W., & Wegner, T.G. (1986). Air Force Officer Qualifying Test Form O: Development and standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Sawin, L.L. (1987). The validity of the AFOQT Pilot composite in predicting UPT outcome for FSP attendees. Unpublished manuscript.

Shanahan, F.M., & Kantor, J.E. (1986). Basic navigator battery: An experimental selection composite for undergraduate navigator training (AFHRL-TR-86-3, AD-A168 857). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Skinner, J., & Ree, M.J. (1987). Air Force Officer Qualifying Test: Item and factor analysis of Form O (AFHRL-TR-86-68, AD-A184 975). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Sperl, T.C. (1988). The development of Quick Score Composites for the Air Force Officer Qualifying Test. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Sperl, T.C., & Ree, M.J. (1989, August). Air Force Officer Qualifying Test (AFOQT): Development of Quick Score Composites for Forms P1 and P2. Paper presented at the 97th Annual Conference of the American Psychological Association, New Orleans, LA.

Steuck, K.W., Watson, T.W., & Skinner, J. (1988). Air Force Officer Qualifying Test (AFOQT): Forms P pre-implementation analyses and equating (AFHRL-TP-88-6, AD-A201 100). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Valentine, L.D., Jr. (1961). Development of the Air Force Precommission Screening Test - 62 (ASD-TN-61-146, AD-269 527). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Valentine, L.D., Jr. (1977). Navigator-observer selection research: Development of a new Air Force Officer Qualifying Test navigator-technical composite (AFHRL-TR-77-36, AD-A042 689). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Valentine, L.D., Jr., & Creager, J.A. (1961). Officer selection and classification tests: Their development and use (ASD-TN-61-145, AD-269 827). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Wherry, R.J., & Gaylord, R.H. (1943). The concept of test and item reliability in relation to factor pattern. Psychometrika, 8, 247-264.

APPENDIX A: Content of Forms M, N, and O of the AFOQT

Table A-1. Content of AFOQT Form M (1975)

Subtest[a]	Items	Composites[b] P	N-T	OQ	V	Q
Booklet 1 (AFPT 972)						
Quantitative Aptitude	60		X	X		X
Booklet 2 (AFPT 973)						
Verbal Aptitude	60			X	X	
Officer Biographical Inventory[c]	96			X		
Booklet 3 (AFPT 974)						
Scale Reading [d]	48		X			
Aerial Landmarks[d]	40		X			
General Science	24		X			
Booklet 4 (AFPT 975)						
Mechanical Information	24	X	X			
Mechanical Principles	24	X	X			
Booklet 5 (AFPT 976)						
Pilot Biographical Inventory	50	X				
Aviation Information	24	X				
Visualization of Maneuvers[d]	24	X				
Instrument Comprehension[d]	24	X				
Stick and Rudder Orientation[d]	24	X				
Total	522					

[a]Scale Reading and Aerial Landmarks are scored according to the number right minus one-fourth the number wrong; Visualization of Maneuvers and Instrument Comprehension are scored as the number right minus one-third the number wrong. Other subtests are scored as number right only.

[b]Composites: Pilot (P), Navigator-Technical (N-T), Officer Quality (OQ), Verbal (V), Quantitative (Q).

[c]Not administered to female applicants.

[d]Speeded subtests.

Table A-2. Content of AFOQT Form N (1978)

Subtests[a]	Items	P	N-T	OQ	V	Q
		Composites[b]				
Booklet 1 (AFPT 982)						
Arithmetic Reasoning	25		X	X		X
Math Knowledge	25		X	X		X
Data Interpretation	25		X	X		X
Booklet 2 (AFPT 983)						
Word Knowledge	25			X	X	
Reading Comprehension	25			X	X	
Background for Current Events	25			X	X	
Verbal Analogies	25	X		X	X	
Booklet 3 (AFPT 984)						
Table Reading[c]	50	X	X			
Electrical Maze[c]	30	X	X			
Block Counting[c]	80	X	X			
Scale Reading[c]	48	X	X			
Tools	25	X	X			
Mechanical Comprehension	24	X	X			
Booklet 4 (AFPT 985)						
Rotated Blocks	20		X			
Aerial Landmarks[c]	40		X			
General Science	24		X			
Instrument Comprehension[c]	24	X				
Pilot Biographic and Attitude Scale	66	X				
Total	606					

[a]Instrument Comprehension is scored according to the number right minus one-third the number wrong; the remaining speeded subtests are scored as the number right minus one-fourth the number wrong. Other subtests are scored as number right only.

[b]Composites: Pilot (P), Navigator-Technical (N-T), Officer Quality (OQ), Verbal (V), Quantitative (Q).

[c]Speeded subtests.

Table A-3. Content of AFOQT Form O (1981)

Subtest[a]	Items	P	N-T	AA	V	Q
		Composites[b]				
Verbal Analogies	25	X		X	X	
Arithmetic Reasoning	25		X	X		X
Reading Comprehension	25			X	X	
Data Interpretation	25		X	X		X
Word Knowledge	25			X	X	
Math Knowledge	25		X	X		X
Mechanical Comprehension	20	X	X			
Electrical Maze	20	X	X			
Scale Reading	40	X	X			
Instrument Comprehension	20	X				
Block Counting	20	X	X			
Table Reading	40	X	X			
Aviation Information	20	X				
Rotated Blocks	15		X			
General Science	20		X			
Hidden Figures	15		X			
Total	380					

[a]All subtests are scored as number right only. No subtests are specifically designated as speeded because all subtests contain elements of both power and speed.

[b]Composites: Pilot (P), Navigator-Technical (N-T), Academic Aptitude (AA), Verbal (V), Quantitative (Q).

APPENDIX B: United States Air Force Officer Qualifying Test Information Pamphlet: Subtest Sample Items

PART 1

Verbal Analogies

DIRECTIONS: This part of the test measures your ability to reason and see relationships between words. You are to choose the answer that best completes the analogy developed at the beginning of each question.

1. FINGER is to HAND as TOOTH is to

1-A	tongue.
1-B	lips.
1-C	nose.
1-D	mouth.
1-E	molar.

2. RACQUET is to COURT as

2-A	tractor is to field.
2-B	blossom is to bloom.
2-C	stalk is to prey.
2-D	plan is to strategy.
2-E	moon is to planet.

3. SWEATER is to CLOTHES as

3-A	bottle is to cork.
3-B	hand is to finger.
3-C	shoe is to foot.
3-D	rose is to flowers.
3-E	dog is to cat.

4. ROW is to BOAT as SAIL is to

4-A	ocean.
4-B	navigate.
4-C	rudder.
4-D	ship.
4-E	travel.

5. FLY is to AIRPLANE as

5-A	drive is to stake.
5-B	skate is to slide.
5-C	push is to fall.
5-D	swim is to float.
5-E	rod is to hook

PART 2

Arithmetic Reasoning

DIRECTIONS: This part of the test measures mathematical reasoning. It is concerned with your ability to arrive at solutions to problems. Each problem is followed by five possible answers. Decide which one of the five answers is most nearly correct.

1. A field with an area of 420 square yards is twice as large in area as a second field. If the second field is 15 yards long, how wide is it?

1-A	7 yards
1-B	14 yards
1-C	28 yards
1-D	56 yards
1-E	90 yards

2. A passenger plane can carry two tons of cargo. A freight plane can carry five tons of cargo. If an equal number of both kinds of planes are used to ship 105 tons of cargo and each plane carries its maximum cargo load, how many tons of cargo are shipped on the passenger planes?

2-A	15.0 tons
2-B	30.0 tons
2-C	42.0 tons
2-D	52.5 tons
2-E	75.0 tons

3. A typist took three typing tests. The average typing speed on these three tests was 48 words per minute. If the typist's speed on two of these tests was 52 words per minute, what was the typist's speed on the third test?

3-A	46 words per minute
3-B	44 words per minute
3-C	42 words per minute
3-D	40 words per minute
3-E	38 words per minute

4. An Air Force recruiting station enlisted 450 people. Of these, 28% were under 20 years old and 32% were 20 to 22 years old. How many of the recruits were over 22 years old?

4-A	130
4-B	140
4-C	175
4-D	180
4-E	270

5. If an aircraft travels at 564 miles per hour, how far did the aircraft fly in 900 seconds?

5-A	141 miles
5-B	226 miles
5-C	300 miles
5-D	451 miles
5-E	846 miles

PART 3

Reading Comprehension

DIRECTIONS: This part of the test measures your ability to read and understand paragraphs. For each question, you are to choose the answer that best completes the meaning of the paragraph.

1. If they are to function effectively, organizations, like other systems, must achieve a natural harmony or coherence among their component parts. The structural and situational elements of an effective organization form themselves into a tightly knit, highly cohesive package. An organization whose parts are mismatched, however, cannot carry out its missions. If managers are to design effective organizations, they need to

   1-A simplify organizational structures.
   1-B encourage greater specialization of labor.
   1-C emphasize the fit of organizational parts.
   1-D introduce more technological innovations.
   1-E reduce the span of control in the organization.

2. First, Clostridium Botulinum, the bacterium that produces the poison, must be present. These bacteria are widespread in the environment and are considered by some to be everywhere, in fact. Second, the bacterium that produces the deadly toxin must be treated to an atmosphere that's free of oxygen and to temperatures that are just warm enough but not too warm. Those conditions have to be held long enough for the toxin to develop. Acid will prevent the growth of the organism and the production of toxin. The following condition is necessary for *Botulism* to develop:

   2-A presence of oxygen.
   2-B a brief period of time.
   2-C presence of acid.
   2-D *warm temperatures.*
   2-E exposure to rare bacteria.

3. Due to our short life span of seventy-odd years it is easy for human beings to think of earth as a planet which never changes. Yet we live on a dynamic planet with many factors contributing to change. We know that wind and rain erode and shape our planet. Many other forces are also at work, such as volcanic activity, temperature fluctuations, and even extraterrestrial interaction such as meteors and gravitational forces. The earth, in actuality, is a large rock

   3-A in a state of inertia.
   3-B which is quickly eroding.
   3-C which is evolving.
   3-D which is subject to temperature fluctuations caused by interplanetary interaction.
   3-E which is subject to winds caused by meteor activity.

4. Mustela nigripes, the rarely seen black-footed ferret, is often confused with Mustela putorius, the common European polecat. It is true that these two mammals resemble each other in some ways. However, they are two distinct and separate species with differences in color, body form, and other attributes. Who knows how many sightings of the black-footed ferret

4–A were the result of seeing the European polecat running loose?
4–B were of species other than the common European polecat?
4–C were made of a related species of the same form and color?
4–D were instead sightings of the Mustela nigripes?
4–E were due to the European polecat destroying their habitat?

5. One theory that explains the similarities between Mayan art and ancient Chinese art is called "diffusion." This theory evolves from the belief that invention is so unique that it happens only once, then is "diffused" to other cultures through travel, trade, and war. This theory might explain why

5–A the airplane and birds both have wings.
5–B certain artifacts in Central America resemble those found in Southeast Asia.
5–C most great art comes from Europe, where there is much travel between countries.
5–D rivers in South America and Africa have some similar features.
5–E England, being so remote in the Middle Ages, is the only country to have castles.

PART 4

Data Interpretation

DIRECTIONS: This part of the test measures your ability to interpret data from tables and graphs. Each table and graph is followed by two, three, or four questions pertaining to that table or graph only.

Ration A	Ration B	Ration C	Ration D
4.5	2.8	3.5	6.2
6.5	3.6	7.6	4.6
3.3	4.9	5.3	5.2
4.9	7.1	5.5	6.1
5.8	3.7	4.1	7.0
2.5	5.1	6.7	3.9
4.2	6.0	4.5	5.9
4.7	4.7	6.5	7.1
3.8	6.1	5.7	5.9
5.3	5.0	5.9	8.1
45.5	49.0	55.3	60.0

Four groups of ten rabbits were each fed a complete commercial ration from birth until 5 months of age. Each animal was weighed at birth and at five months. The weight of the animal at birth was subtracted from its weight at five months. These differences, in pounds, are shown above for each rabbit.

1. Which ration appears to be the best?

   1-A Ration A
   1-B Ration B
   1-C Ration C
   1-D Ration D

2. The average weight gain of the rabbits fed Ration B is

   2-A 4.55 lbs.
   2-B 4.90 lbs.
   2-C 5.53 lbs.
   2-D 5.44 lbs.
   2-E 6.00 lbs.

3. The average difference in weight gain between the group of rabbits who gained the most and the group who gained the least was

3-A 0.35 lbs.
3-B 0.45 lbs.
3-C 0.47 lbs.
3-D 1.10 lbs.
3-E 1.45 lbs.

4. The individual rabbit showing the least gain was fed

4-A Ration A.
4-B Ration B.
4-C Ration C.
4-D Ration D.

5. The individual rabbit that gained the most weight gained

5-A 6.0 lbs.
5-B 6.5 lbs.
5-C 7.1 lbs.
5-D 7.6 lbs.
5-E 8.1 lbs.

PART 5

Word Knowledge

DIRECTIONS: This part of the test measures verbal comprehension involving your ability to understand written language. For each question, choose the answer that means the same as the capitalized word.

1. CRIMSON

   1-A crisp
   1-B neatly dressed
   1-C reddish
   1-D colorful
   1-E lively

2. CEASE

   2-A start
   2-B change
   2-C continue
   2-D stop
   2-E fold

3. FORTNIGHT

   3-A two days
   3-B one week
   3-C two weeks
   3-D one month
   3-E two months

4. SULLEN

   4-A grayish yellow
   4-B soaking wet
   4-C very dirty
   4-D angrily silent
   4-E mildly nauseated

5. TERSE

   5-A pointed
   5-B trivial
   5-C oral
   5-D lengthy
   5-E raggedy

PART 6

Math Knowledge

DIRECTIONS: This part of the test measures your ability to use learned mathematical relationships. Each problem is followed by five possible answers. Decide which one of the five answers is most nearly correct.

1. The first digit of the square root of 59043 is

   1-A 1
   1-B 2
   1-C 3
   1-D 4
   1-E 5

2. The distance in miles around a circular course with a radius of 35 miles is (use pi = 22/7)

   2-A 110 miles.
   2-B 156 miles.
   2-C 220 miles.
   2-D 440 miles.
   2-E 880 miles.

3. The expression "3 factorial" equals

   3-A 1 9
   3-B 1 6
   3-C 6
   3-D 9
   3-E 27

4. Solve for x: $2x\ 7 = 2x^2$

   4-A 1 7
   4-B 2 7
   4-C 2
   4-D 7
   4-E 14

5. The reciprocal of 5 is

   5-A 0.1.
   5-B 0.2.
   5-C 0.5.
   5-D 1.0.
   5-E 2.0.

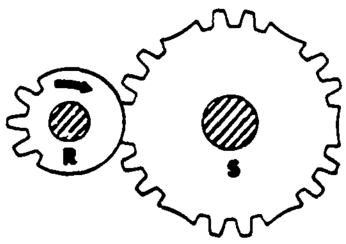
PART 7

Mechanical Comprehension

DIRECTIONS: This part of the test measures your ability to learn and reason with mechanical terms. Choose the answer that best completes the statement. Also included in this part of the test are diagrams of mechanical devices. Following each diagram are several questions or incomplete statements. Study the diagram carefully and select the choice that best answers the question or completes the statement.

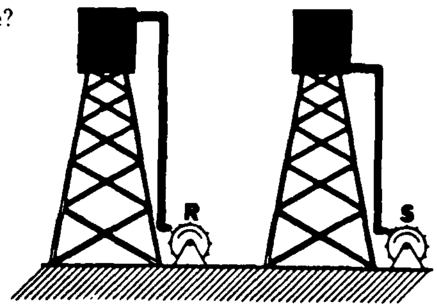
1. If gear R is the driver, at the moment shown, gear S is

   1-A not moving.
   1-B jammed.
   1-C moving at its highest speed.
   1-D moving in the same direction as gear R.
   1-E moving in the opposite direction as gear R.



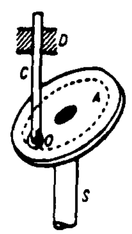
2. Which water wheel will turn for the longer time?

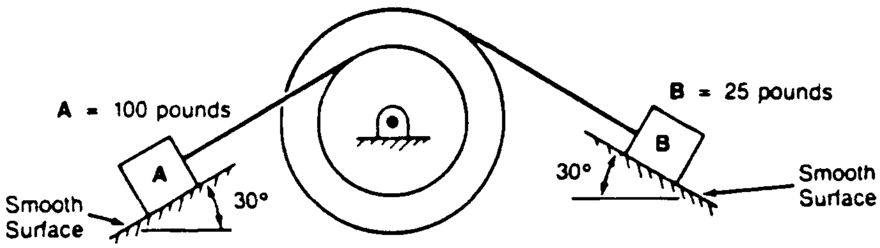
   2-A R
   2-B S
   2-C Both wheels will turn for an equal amount of time.
   2-D Neither wheel will turn at all.
   2-E This can't be determined from the drawing.



3. As shaft S makes one complete turn from the position shown, C moves

   3-A left and then right.
   3-B right and then left.
   3-C up only.
   3-D down only.
   3-E up and down.

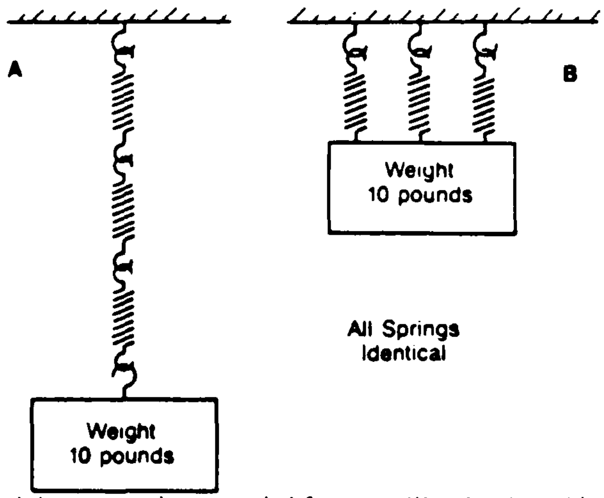




(NOT IN EQUILIBRIUM)

4. If weight B is to slide to the right, what change must be made in the diagram?

4–A	The slope of the inclined plane under A must be increased.
4–B	The slope of the inclined plane under B must be increased.
4–C	The radius of the inner pulley must be decreased.
4–D	The radius of the inner pulley must be increased to a size nearer to that of the outer pulley.
4–E	The radius of the outer pulley must be twice that of the inner pulley.



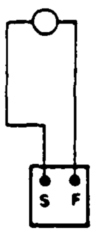
5. Ten-pound weights are each suspended from a ceiling by three identical springs. In A, the extension of each spring is

5–A	nine times greater than in B.
5–B	three times greater than in B.
5–C	the same as in B.
5–D	1/3 less than in B.
5–E	1/9 less than in B.

PART 8

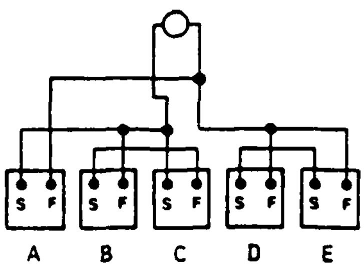
Electrical Maze

DIRECTIONS: This is a test of your ability to choose a correct path from among several choices. In the picture below is a box with dots marked S and F. S is the starting point, and F is the finishing point. You are to follow the line from S, through the circle at the top of the picture, and back to F.



In the problems in this test, there will be five such boxes. Only *one* box will have a line from the S, through the circle, and back to the F in the same box. Dots on the lines show the *only* places where turns or direction changes can be made between lines. If lines meet or cross where there is *no dot*, turns or direction changes *cannot be made*. Now try sample problem S1.

S1.

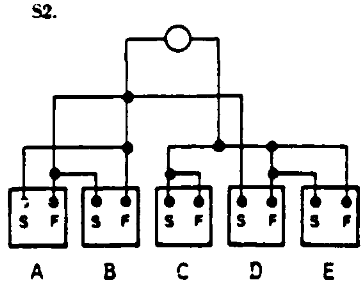


The first box is the one which has the line from S, through the circle, and back to F. Therefore, A is the right answer.

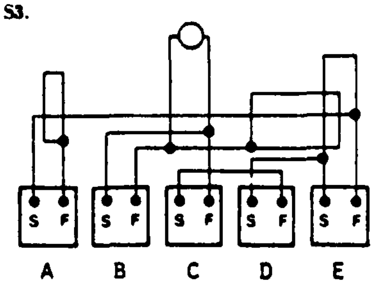
Each diagram in the test has only one box which has a line through the circle and back to F. Some lines are wrong because they lead to a dead end. Some lines are wrong because they come back to the box without going through the circle. Some lines are wrong because they lead to other boxes. Some are wrong because they retrace the same line.

Now try sample problems S2 and S3.

S2.

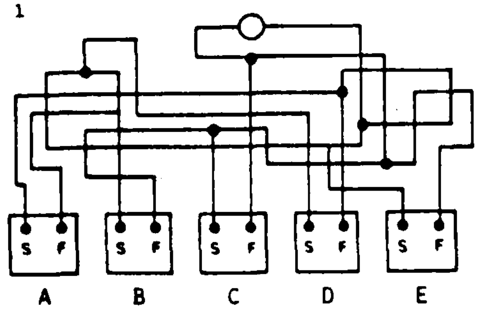


S3.

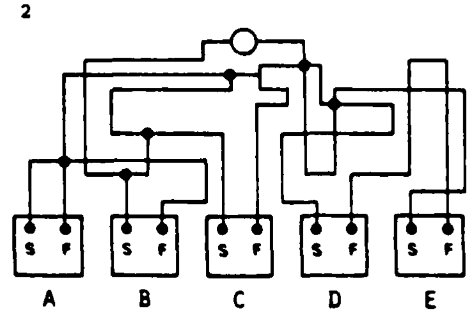


For sample problem S2, the correct answer is D. For sample problem S3, the correct answer is B.

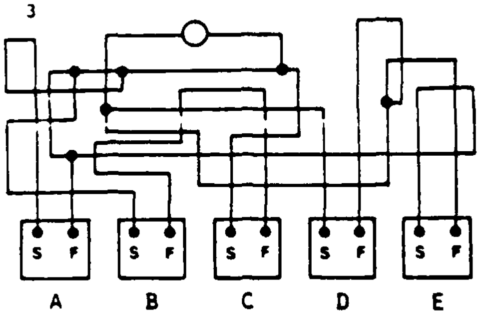
1



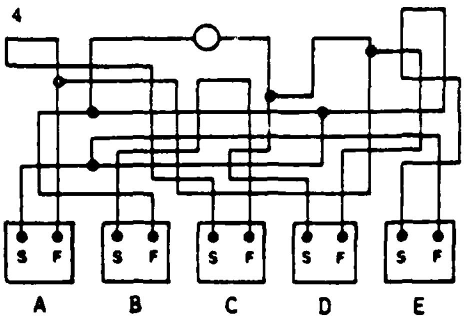
2



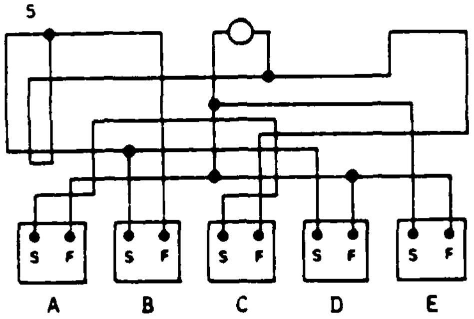
3



4



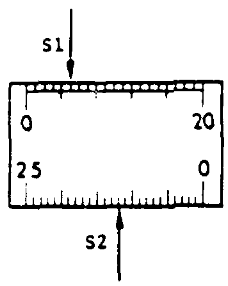
5



PART 9

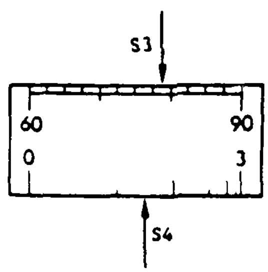
Scale Reading

DIRECTIONS: This is a test of your ability to read scales, dials, and meters. You will be given a variety of scales with various points indicated on them by numbered arrows. You are to estimate the numerical value indicated by each arrow, find the choice closest to this value in the item having the same numbers as the arrow, and then mark your answer on a separate sheet of paper. Now look at the sample items below.



S1. A 6.00
B 5.00
C 4.25
D 2.25
E 1.25

S2. A 13.0
B 12.0
C 10.2
D 1.3
E 1.2



S3. A 81.75
B 79.5
C 78.75
D 77.60
E 67.50

S4. A 1.75
B 1.65
C 1.50
D .75
E .65

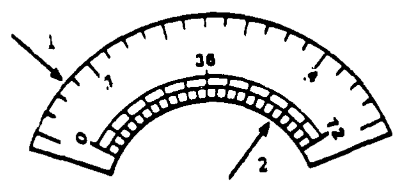
In sample item S1 there are five subdivisions of four steps each between 0 and 20. The arrow points between the long subdivision markers representing 4 and 8. Since it points to the marker that is one step to the right of subdivision marker 4, it points to 5.00. This is choice B in sample item S1.

In sample item S2 the scale runs from right to left. There are five subdivisions of five steps each, so each step represents .1, and the arrow points to the marker representing 1.2. This is choice E in sample item S2.

In sample item S3 the arrow points between two markers. You must estimate the fractional part of the step as accurately as possible. Since the arrow points halfway between the markers representing 77.5 and 80.0, it points to 78.75. This is choice C in sample item S3.

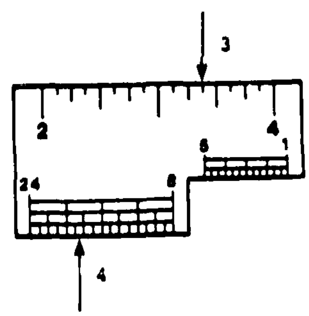
In sample item S4 each step represents .5, but the steps are of unequal width with each step being two-thirds as wide as the preceding one. Therefore, the scale is compressed as the values increase. The arrow is pointing to a position halfway between the marker representing .5 and 1.0, but because of the compression of the scale the value of this point must be less than .75. Actually, it is .65, which is choice E in sample item S4.

1. A 735
   B 725
   C 680
   D .660
   E .570



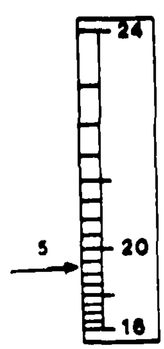
2. A 86
   B 81
   C 64
   D 60
   E 58

3. A 3.425
   B 3.375
   C 3.275
   D 3.150
   E 3.125



4. A 22.6
   B 21.2
   C 18.5
   D 17.5
   E 13.2

5. A 20.7
   B 20.3
   C 19.9
   D 19.3
   E 19.2



PART 10

Instrument Comprehension

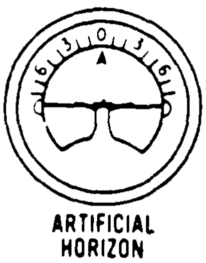
DIRECTIONS: This test measures your ability to determine the position of an airplane in flight from reading instruments showing its compass heading, its amount of climb or dive, and its degree of bank to right or left. In each item the left-hand dial is labeled ARTIFICIAL HORIZON. On the face of this dial the small aircraft silhouette remains stationary, while the positions of the heavy black line and the black pointer vary with changes in the position of the airplane in which the instrument is located.

The heavy black line represents the HORIZON LINE, and the black pointer shows the degree of BANK to right or left.

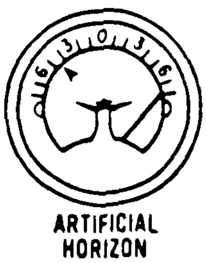
If the airplane is neither climbing nor diving, the horizon line is directly on the silhouette's fuselage, as in dial 1, below.

If the airplane is climbing, the fuselage silhouette is seen between the horizon line and the pointer, as in dial 2, below. The greater the amount of climb, the greater the distance between the horizon line and the fuselage silhouette.

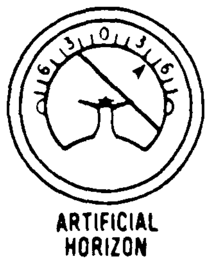
If the airplane is diving, the horizon line is seen between the fuselage silhouette and the pointer, as in dial 3, below. The greater the amount of dive, the greater the distance between the horizon line and the fuselage silhouette.



Dial 1



Dial 2



Dial 3

If the airplane has no bank, the black pointer is seen to point to zero, as in dial 1, above.

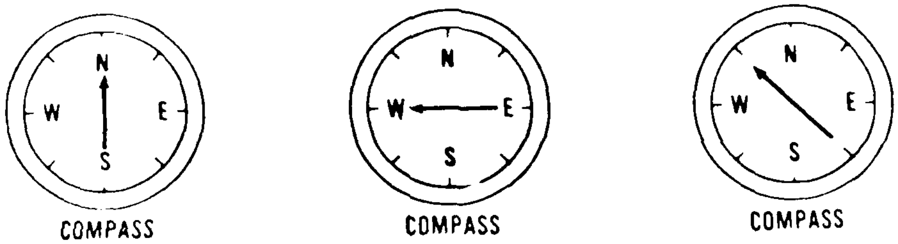
If the airplane is banked to the pilot's right, the pointer is seen to the left of zero, as in dial 2, above.

If the airplane is banked to the pilot's left, the pointer is seen to the right of zero, as in dial 3, above.

The HORIZON LINE tilts as the aircraft is banked and is always at right angles to the pointer.

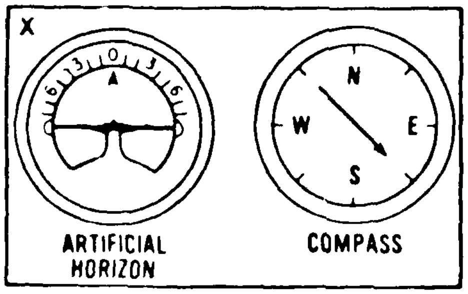
Dial 1 above shows an airplane neither climbing nor diving, with no bank.
Dial 2 above shows an airplane climbing and banking 45 degrees to the pilot's right.
Dial 3 above shows an airplane diving and banked 45 degrees to the pilot's left.

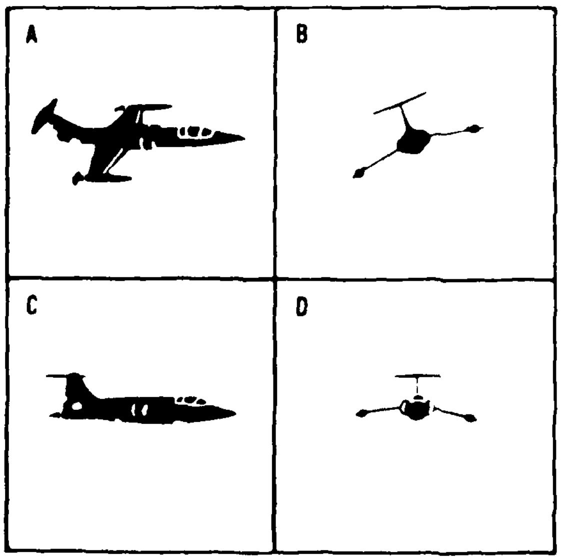
On each item the right-hand dial is labeled COMPASS. On this dial, the arrow shows the compass direction in which the airplane is headed at the moment. Dial 4 shows it headed north; dial 5 shows it headed west; and dial 6 shows it headed northwest.



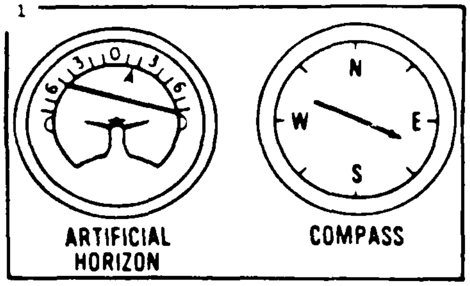
Dial 4 Dial 5 Dial 6

Each item in this test consists of two dials and four silhouettes of airplanes in flight. Your task is to determine which one of the four airplanes is MOST NEARLY in the position indicated by the two dials. YOU ARE ALWAYS LOOKING NORTH AT THE SAME ALTITUDE AS EACH OF THE PLANES. EAST IS ALWAYS TO YOUR RIGHT AS YOU LOOK AT THE PAGE. Item X is a sample. In item X the dial labeled ARTIFICIAL HORIZON shows that the airplane is NOT banked, and is neither climbing nor diving. The COMPASS shows that it is headed southeast. The only one of the four airplane silhouettes that meets these specifications is in the box lettered C, so the answer to X is C. Note that B is a rear view, while D is a front view. Note also that A is banked to the right and that B is banked to the left.

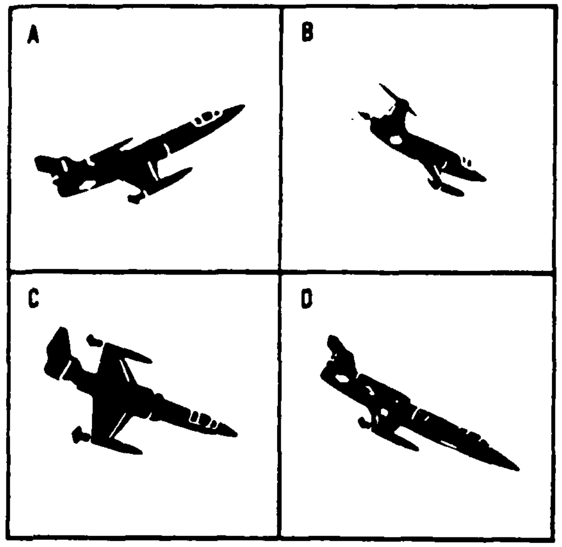




1

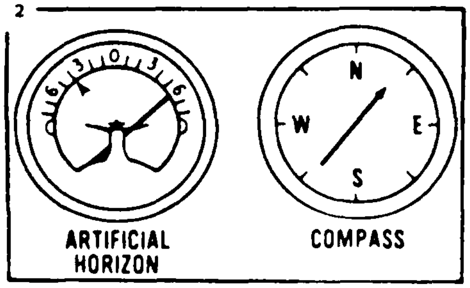


6
3
0
3
6
N
W
E
S
ARTIFICIAL HORIZON
COMPASS

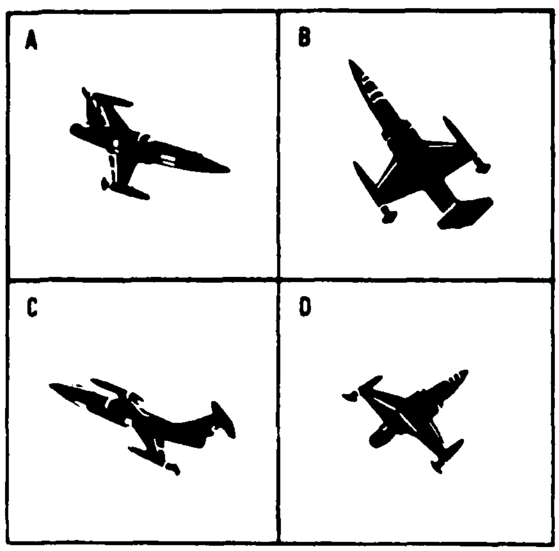


A
B
C
D

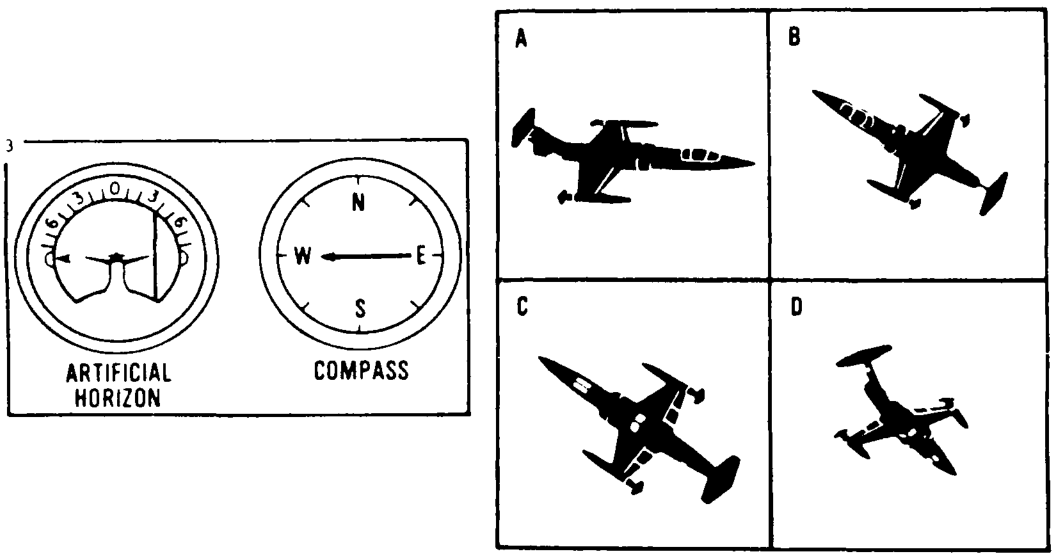
2

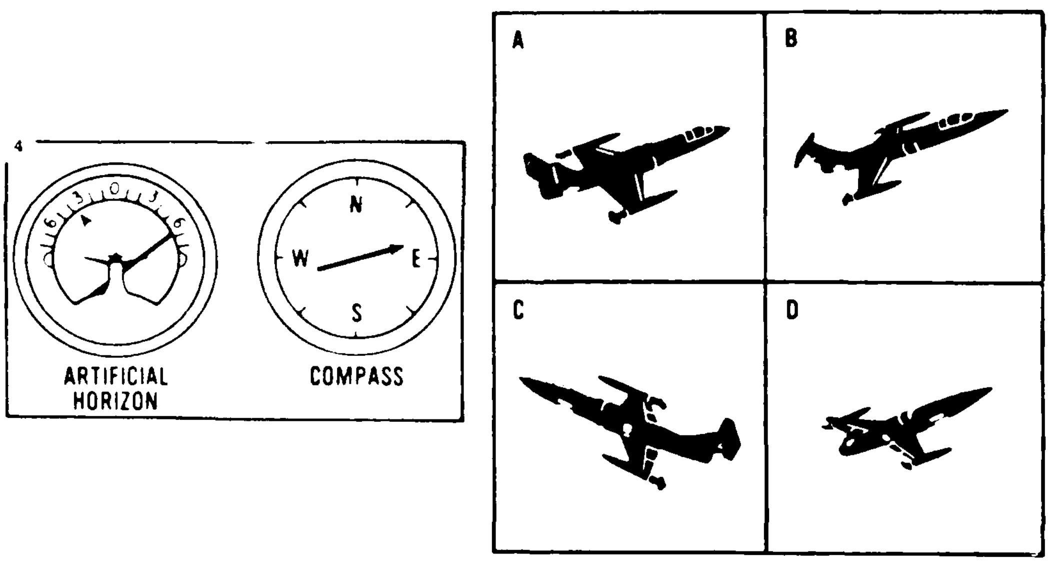


6
3
0
3
6
N
W
E
S
ARTIFICIAL HORIZON
COMPASS

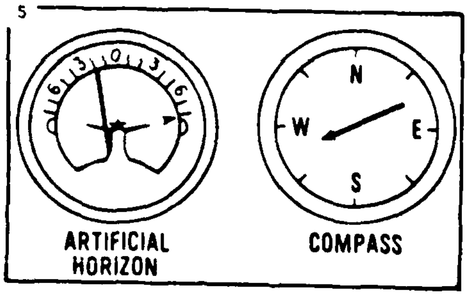


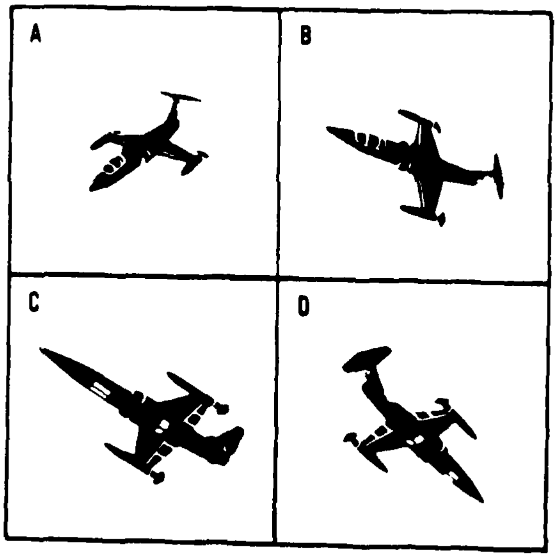
A
B
C
D


3
A
B
C
D
N
W
E
S
ARTIFICIAL HORIZON
COMPASS


4
A
B
C
D
N
W
E
S
ARTIFICIAL HORIZON
COMPASS

5

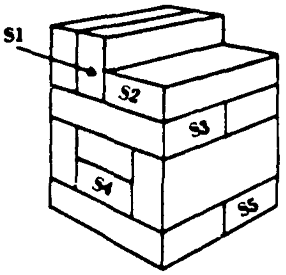




PART 11

Block Counting

DIRECTIONS: This is a test of your ability to "see into" a 3-dimensional pile of blocks and determine how many pieces are touched by certain numbered blocks. *All of the blocks in each pile are the same size and shape.* Look at the sample below.

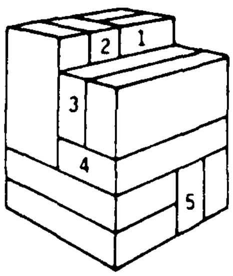


KEY					
Block	A	B	C	D	E
S1	1	2	3	4	5
S2	3	4	5	6	7
S3	5	6	7	8	9
S4	2	3	4	5	6
S5	2	3	4	5	6

Block S1 touches the other 2 top blocks and the 2 blocks directly below it. The total number of blocks touched by S1 is, therefore, 4. For sample problem S1, 4 is choice D in the key to the right.

Block S2 touches blocks 1 and 3 and the unlettered block to the right of block 3. Since block S2 touches 3 other blocks, the answer is 3. According to the key, 3 is choice A for sample problem S2. Now look at sample problem S3. It touches 3 blocks above, 3 blocks below, and one block on the right. Therefore, the correct answer is 7, so C is the correct answer to sample problem S3.

Now count the blocks touching blocks S4 and S5. For block S4, the correct answer is 5, so D would be the correct answer. For block S5, the correct answer is 4, so C would be the correct answer.



KEY					
Block	**A**	**B**	**C**	**D**	**E**
1	3	4	5	6	7
2	5	6	7	8	9
3	5	6	7	8	9
4	4	5	6	7	8
5	4	5	6	7	8

PART 12

Table Reading

DIRECTIONS: This is a test of your ability to read tables quickly and accurately. Look at the table below. Notice that the X values are shown at the top of the table and the Y values are shown on the left of the table. In this test, you are to find the entry that occurs at the intersection of the row and the column corresponding to the values given.

X VALUE

Y VALUE	-3	-2	-1	0	+1	+2	+3
+3	22	23	25	27	28	29	30
+2	23	25	27	29	30	31	32
+1	24	26	28	30	32	33	34
0	26	27	29	31	33	34	35
-1	27	29	30	32	34	35	37
-2	28	30	31	33	35	36	38
-3	29	31	32	34	36	37	39

	X	Y	A	B	C	D	E
1.	0	-1	29	33	32	35	34
2.	-3	-3	22	29	23	31	28
3.	-1	+2	25	31	29	30	27
4.	+3	0	30	34	35	37	39
5.	-2	+1	26	23	29	25	22

PART 13

Aviation Information

DIRECTIONS: This part of the test measures your knowledge of aviation. Each of the questions or incomplete statements is followed by five choices. Decide which one of the choices best answers the question or completes the statement.

1. If the elevator tabs on a plane are lowered, the plane will tend to

1-A	nose up.
1-B	nose down.
1-C	pitch fore and aft.
1-D	go into a slow roll.
1-E	wing over.

2. The pilot always advances the throttle during a

2-A	nose dive.
2-B	landing.
2-C	turn.
2-D	spin.
2-E	climb.

3. The pilot of an airplane can best detect the approach of a stall by the

3-A	increase in speed of the engine.
3-B	increase in pitch and intensity of the sound of the air moving past the plane.
3-C	increase in effectiveness of the rudder.
3-D	ineffectiveness of the ailerons and elevator.
3-E	decrease in pitch and intensity of the sound of the air moving past the plane.

4. It is ordinarily desirable to provide an unusually long flight strip at municipal airports for the takeoff of

4-A	military planes in echelon.
4-B	heavily loaded ships in still air.
4-C	small airplanes in rainy weather.
4-D	any airplane across the wind.
4-E	airplanes that have high climbing speeds.

5. The slipstream of an airplane will have the least effect on the plane's direction of flight when the plane's

5-A	stick is moved forward.
5-B	stick is moved backward.
5-C	stick is moved sideways.
5-D	rudder bar is pressed with the left foot.
5-E	rudder bar is pressed with the right foot.

PART 14

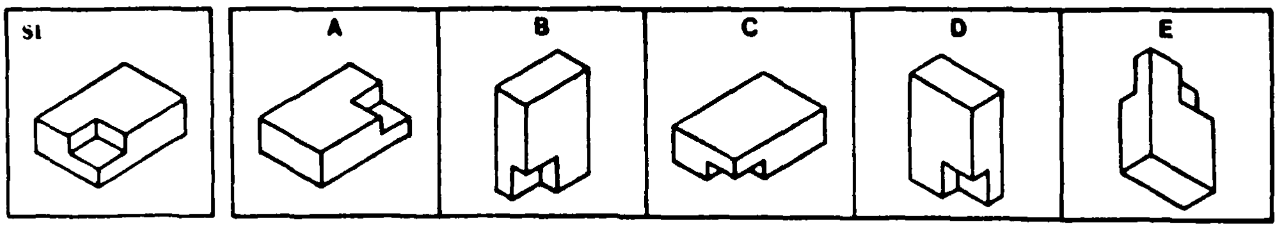
Rotated Blocks

DIRECTIONS: This test measures your ability to visualize and manipulate objects in space. In each item of this test you are shown a picture of a block. The problem is to find a second block which is just like the first.

Look at the two blocks below. Although you see them from different points, the blocks are just alike.

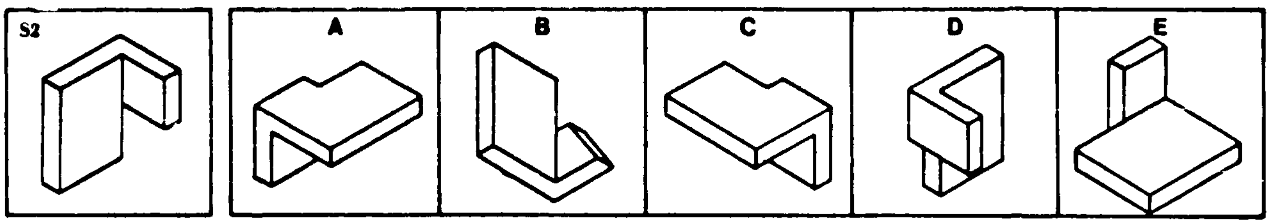
Look at the two blocks below. They are not alike. They can never be turned so that they will be alike.

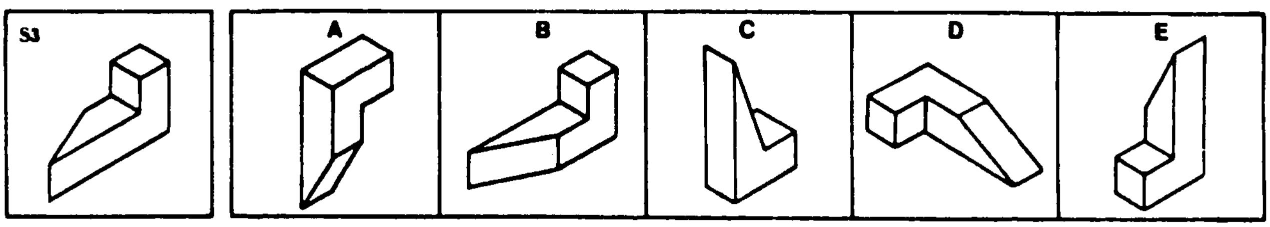
Now look at the sample item below. Which of the five choices is just like the first block?



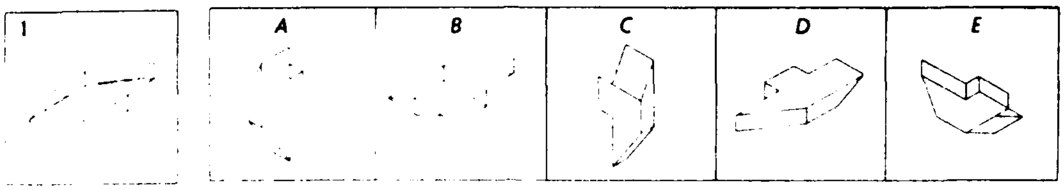
The correct answer is D. It is the same block as seen from a different side.

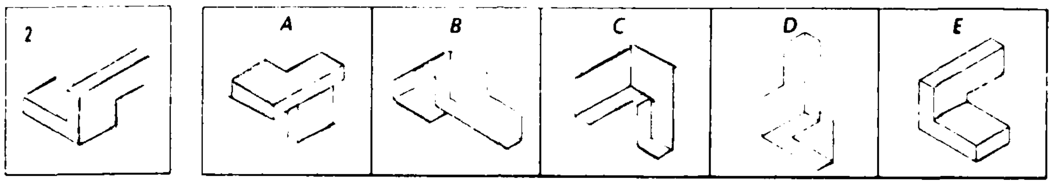
Now look at the two sample items below.

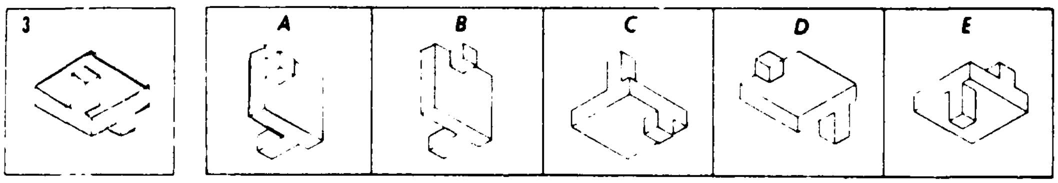


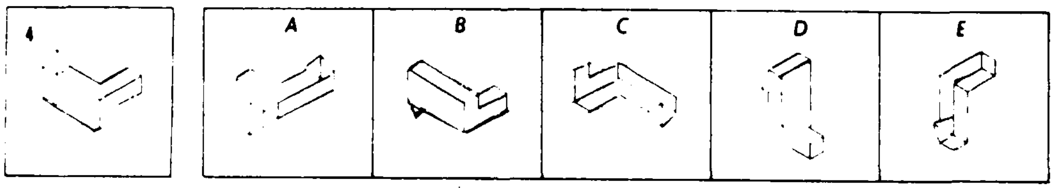


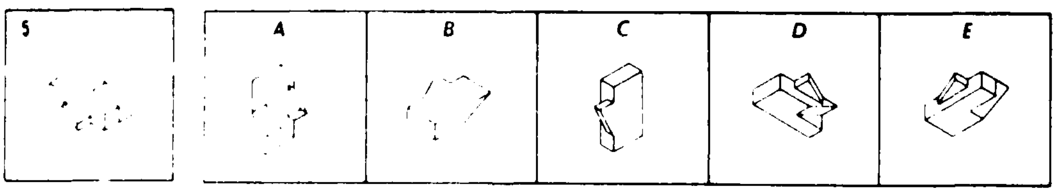
The right answer for S2 is C. The right answer for S3 is A.


1
A
B
C
D
E


2
A
B
C
D
E


3
A
B
C
D
E


4
A
B
C
D
E


5
A
B
C
D
E

PART 15

General Science

DIRECTIONS: This part of the test measures your knowledge in the arena of science. Each of the questions or incomplete statements is followed by five choices. Decide which one of the choices best answers the question or completes the statement.

1. An eclipse of the sun throws the shadow of the

   1-A moon on the sun.
   1-B earth on the sun.
   1-C sun on the earth.
   1-D earth on the moon.
   1-E moon on the earth.

2. Substances which hasten a chemical reaction without themselves undergoing change are called

   2-A buffers.
   2-B catalysts.
   2-C colloids.
   2-D reducers.
   2-E polymers.

3. Lack of iodine is often related to which of the following diseases?

   3-A Beriberi
   3-B Scurvy
   3-C Rickets
   3-D Goiter
   3-E Asthma

4. The chief nutrient in lean meat is

   4-A starch.
   4-B protein.
   4-C fat.
   4-D carbohydrates.
   4-E Vitamin B.

5. After adding salt to water, the freezing point of the water is

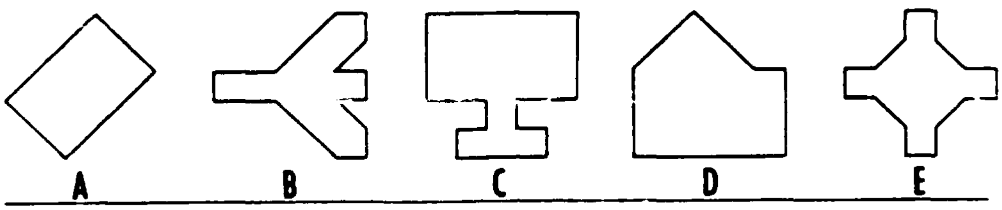
   5-A variable.
   5-B inverted.
   5-C the same.
   5-D raised.
   5-E lowered.

PART 16

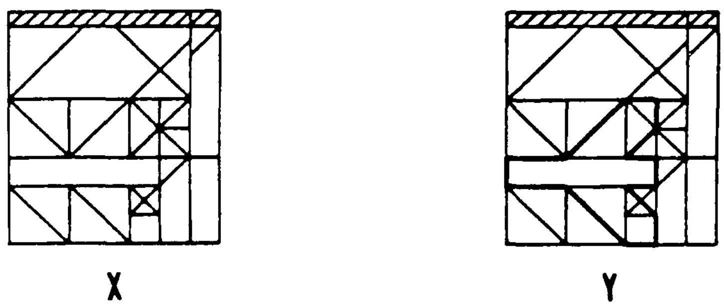
Hidden Figures

DIRECTIONS: This part of the test measures your ability to see a simple figure in a complex drawing. At the top of each page are five figures, lettered A, B, C, D, and E. Below these on each page are several numbered drawings. You are to determine which lettered figure is contained in each of the numbered drawings.

The lettered figures are:



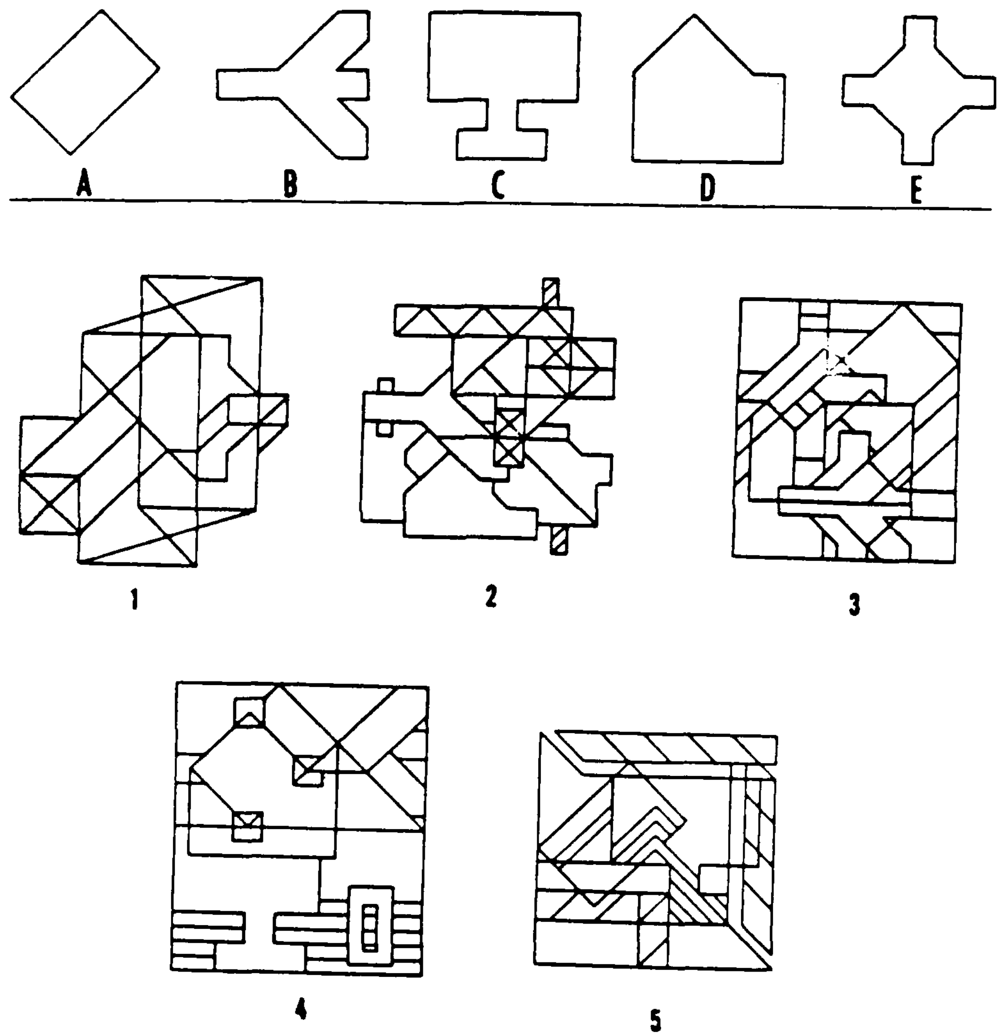
As an example, look at drawing X below.



Which one of the five figures is contained in drawing X?

Now look at drawing Y, which is exactly like drawing X except that the outline of figure B has been made heavy to show where to look for it. Thus, B is the answer to sample item X.

Each numbered drawing contains only *one* of the lettered figures. The correct figure in each drawing will always be of the same size and in the same position as it appears at the top of the page. Therefore, do not rotate the page in order to find it. Look at each numbered drawing and decide which one of the five lettered figures is contained in it.


A
B
C
D
E
1
2
3
4
5

Sample Item Key

Verbal Analogies

1. D
2. A
3. D
4. D
5. A

Arithmetic Reasoning

1. B
2. B
3. D
4. D
5. A

Reading Comprehension

1. C
2. D
3. C
4. A
5. B

Data Interpretation

1. D
2. B
3. E
4. A
5. E

Word Knowledge

1. C
2. D
3. C
4. D
5. A

Math Knowledge

1. B
2. C
3. C
4. A
5. B

Mechanical Comprehension

1. A
2. B
3. E
4. C
5. B

Electrical Maze

1. B
2. C
3. E
4. A
5. D

Scale Reading

1. A
2. E
3. B
4. C
5. D

Instrument Comprehension

1. D
2. D
3. C
4. A
5. B

Block Counting

1. D
2. B
3. A
4. E
5. C

Table Reading

1. C
2. B
3. E
4. C
5. A

Aviation Information

1. B
2. E
3. D
4. B
5. C

Rotated Blocks

1. D
2. C
3. C
4. D
5. B

General Science

1. E
2. B
3. D
4. B
5. E

Hidden Figures

1. A
2. A
3. E
4. D
5. C

APPENDIX C: Skew and Kurtosis of AFOQT Form P Subtest and Composite Raw Scores

Table C-1. Skew and Kurtosis of AFOQT Form P Subtest and Composite Raw Scores

Subtest	Number of items scored	Skew P1	Skew P2	Kurtosis P1	Kurtosis P2
Verbal Analogies	25	-.41	-.58	-.32	-.07
Arithmetic Reasoning	25	-.25	-.20	-.83	-.81
Reading Comprehension	25	-.24	-.39	-.57	-.64
Data Interpretation	25	-.27	-.42	-.57	-.56
Word Knowledge	25	-.18	-.12	-.90	-.98
Math Knowledge	25	-.42	-.14	-.99	-1.10
Mechanical Comprehension	20	.11	.09	-.72	-.73
Electrical Maze	20	.46	.37	.08	.01
Scale Reading	40	-.09	-.09	-.58	-.32
Instrument Comprehension	20	-.13	-.25	-1.13	-1.155
Block Counting	20	-.43	-.28	-.35	-.37
Table Reading	40	-.53	-.40	.43	.10
Aviation Information	19	.54	.54	-.54	-.54
Rotated Blocks	15	-.07	-.23	-.62	-.73
General Science	19	.17	.21	-.72	-.64
Hidden Figures	15	-.20	-.24	-.71	-.51
Composite					
Pilot	204	-.27	-.34	-.40	-.30
Navigator-Technical	264	-.32	-.31	-.40	-.39
Academic Aptitude	150	-.35	-.34	-.50	-.56
Verbal	75	-.27	-.35	-.58	-.62
Quantitative	75	-.32	-.24	-.74	-.85

APPENDIX D: Item Analysis Summary

Table D-1. Mean Difficulties, Biserials, and Point Biserials, AFOQT Form P1

Subtest		Mean item difficulty	Mean biserial	Mean point biserial
Verbal Analogies	(VA)	.648	.573	.410
Arithmetic Reasoning	(AR)	.619	.667	.500
Reading Comprehension	(RC)	.607	.578	.436
Data Interpretation	(DI)	.640	.517	.390
Word Knowledge	(WK)	.606	.662	.500
Math Knowledge	(MK)	.665	.722	.544
Mechanical Comprehension	(MC)	.496	.545	.429
Electrical Maze	(EM)	.388	.549	.411
Scale Reading	(SR)	.595	.509	.379
Instrument Comprehension	(IC)	.576	.700	.546
Block Counting	(BC)	.640	.663	.488
Table Reading	(TR)	.690	.683	.456
Aviation Information	(AI)	.447	.595	.461
Rotated Blocks	(RB)	.533	.631	.470
General Science	(GS)	.491	.560	.434
Hidden Figures	(HF)	.666	.659	.474

Table D-2. Distribution of Item Difficulty, AFOQT Form P1

Subtest		Number of items with item difficulty of				
		.00 to .20	.21 to .40	.41 to .60	.61 to .80	.81 to .99
Verbal Analogies	(VA)	0	3	8	8	6
Arithmetic Reasoning	(AR)	0	3	8	10	4
Reading Comprehension	(RC)	0	7	3	12	3
Data Interpretation	(DI)	0	1	11	9	4
Word Knowledge	(WK)	0	5	7	10	3
Math Knowledge	(MK)	0	0	8	15	2
Mechanical Comprehension	(MC)	0	3	13	4	0
Electrical Maze	(EM)	6	5	4	5	0
Scale Reading	(SR)	0	9	11	12	8
Instrument Comprehension	(IC)	0	2	9	9	0
Block Counting	(BC)	0	3	6	6	5
Table Reading	(TR)	2	7	5	5	21
Aviation Information	(AI)	0	8	9	1	1
Rotated Blocks	(RB)	0	4	6	2	3
General Science	(GS)	0	6	8	4	1
Hidden Figures	(HF)	0	1	6	4	4

Table D-3. Distribution of Item Discrimination, AFOQT Form P1

Subtest		Number of items with item biserial of .00 to .20	.21 to .40	.41 to .60	.61 to .80	.81 to .99
Verbal Analogies	(VA)	0	0	16	9	0
Arithmetic Reasoning	(AR)	0	0	4	20	1
Reading Comprehension	(RC)	0	1	14	10	0
Data Interpretation	(DI)	0	5	13	7	0
Word Knowledge	(WK)	0	0	3	22	0
Math Knowledge	(MK)	0	1	3	14	7
Mechanical Comprehension	(MC)	0	2	14	4	0
Electrical Maze	(EM)	0	1	13	6	0
Scale Reading	(SR)	0	10	19	11	0
Instrument Comprehension	(IC)	0	0	4	13	3
Block Counting	(BC)	0	0	7	12	1
Table Reading	(TR)	0	2	9	17	12
Aviation Information	(AI)	0	1	8	10	0
Rotated Blocks	(RB)	0	0	6	9	0
General Science	(GS)	0	2	8	9	0
Hidden Figures	(HF)	0	0	3	12	0

Table D-4. Mean Difficulties, Biserials, and Point Biserials, AFOQT Form P2

Subtest		Mean item difficulty	Mean biserial	Mean point biserial
Verbal Analogies	(VA)	.676	.579	.410
Arithmetic Reasoning	(AR)	.604	.648	.486
Reading Comprehension	(RC)	.663	.631	.468
Data Interpretation	(DI)	.667	.574	.425
Word Knowledge	(WK)	.603	.652	.497
Math Knowledge	(MK)	.610	.668	.514
Mechanical Comprehension	(MC)	.498	.552	.432
Electrical Maze	(EM)	.388	.533	.393
Scale Reading	(SR)	.563	.482	.362
Instrument Comprehension	(IC)	.599	.740	.572
Block Counting	(BC)	.585	.606	.447
Table Reading	(TR)	.690	.649	.434
Aviation Information	(AI)	.476	.607	.468
Rotated Blocks	(RB)	.564	.646	.489
General Science	(GS)	.482	.518	.402
Hidden Figures	(HF)	.652	.648	.460

Table D-5. Distribution of Item Difficulty, AFOQT Form P2

Subtest		Number of items with item difficulty of				
		.00 to .20	.21 to .40	.41 to .60	.61 to .80	.81 to .99
Verbal Analogies	(VA)	0	2	8	7	8
Arithmetic Reasoning	(AR)	1	2	9	9	4
Reading Comprehension	(RC)	0	1	9	10	5
Data Interpretation	(DI)	0	1	7	14	3
Word Knowledge	(WK)	0	4	9	10	2
Math Knowledge	(MK)	0	0	15	8	2
Mechanical Comprehension	(MC)	0	4	10	6	0
Electrical Maze	(EM)	8	3	4	5	0
Scale Reading	(SR)	1	12	9	12	6
Instrument Comprehension	(IC)	0	0	11	8	1
Block Counting	(BC)	0	6	4	7	3
Table Reading	(TR)	1	8	4	6	21
Aviation Information	(AI)	0	8	8	2	1
Rotated Blocks	(RB)	0	3	7	2	3
General Science	(GS)	0	6	10	2	1
Hidden Figures	(HF)	0	3	4	2	6

Table D-6. Distribution of Item Discrimination, AFOQT Form P2

Subtest		Number of items with item biserial of .00 to .20	.21 to .40	.41 to .60	.61 to .80	.81 to .99
Verbal Analogies	(VA)	0	0	14	11	0
Arithmetic Reasoning	(AR)	0	0	6	19	0
Reading Comprehension	(RC)	0	0	11	14	0
Data Interpretation	(DI)	0	3	10	12	0
Word Knowledge	(WK)	0	0	6	18	1
Math Knowledge	(MK)	0	0	8	14	3
Mechanical Comprehension	(MC)	0	3	10	7	0
Electrical Maze	(EM)	0	3	12	5	0
Scale Reading	(SR)	0	11	22	7	0
Instrument Comprehension	(IC)	0	0	3	10	7
Block Counting	(BC)	0	2	7	11	0
Table Reading	(TR)	0	3	11	16	10
Aviation Information	(AI)	0	0	7	12	0
Rotated Blocks	(RB)	0	0	4	11	0
General Science	(GS)	0	4	12	3	0
Hidden Figures	(HF)	0	0	3	12	0

APPENDIX E: AFOQT Subtest Intercorrelation matrices
for Forms P1 and P2

Table E-1. Subtest Raw Score Intercorrelations, AFOQT Form P1

	VA	AR	RC	DI	WK	MK	MC	EM	SR	IC	BC	TR	AI	RB	GS
AR	.59														
RC	.67	.55													
DI	.59	.72	.60												
WK	.65	.46	.72	.48											
MK	.54	.72	.46	.58	.36										
MC	.49	.55	.42	.50	.37	.45									
EM	.29	.38	.25	.37	.16	.36	.45								
SR	.46	.67	.43	.65	.32	.55	.47	.46							
IC	.40	.41	.33	.42	.25	.38	.53	.47	.50						
BC	.41	.50	.37	.52	.25	.44	.44	.46	.59	.50					
TR	.31	.40	.34	.46	.23	.38	.26	.32	.53	.35	.50				
AI	.33	.30	.32	.33	.30	.22	.53	.31	.32	.54	.29	.22			
RB	.41	.48	.28	.44	.23	.46	.54	.44	.47	.53	.53	.29	.35		
GS	.55	.55	.52	.49	.49	.55	.61	.39	.43	.45	.38	.22	.46	.45	
HF	.42	.42	.34	.43	.26	.42	.42	.39	.44	.43	.48	.34	.26	.51	.38

Table E-2. Subtest Raw Score Intercorrelations, AFOQT Form P2

	VA	AR	RC	DI	WK	MK	MC	EM	SR	IC	BC	TR	AI	RB	GS
AR	.62														
RC	.72	.56													
DI	.64	.76	.61												
WK	.70	.47	.74	.52											
MK	.53	.74	.46	.64	.37										
MC	.52	.54	.46	.53	.42	.43									
EM	.31	.40	.27	.40	.17	.38	.45								
SR	.45	.63	.42	.61	.31	.53	.43	.45							
IC	.42	.46	.36	.45	.27	.41	.56	.48	.51						
BC	.40	.49	.37	.51	.25	.43	.42	.49	.57	.50					
TR	.32	.44	.31	.45	.21	.37	.28	.34	.55	.37	.51				
AI	.35	.30	.35	.34	.32	.21	.54	.30	.33	.55	.29	.23			
RB	.40	.49	.30	.47	.25	.46	.54	.45	.47	.56	.53	.32	.35		
GS	.56	.55	.53	.52	.54	.51	.65	.40	.42	.49	.37	.24	.49	.46	
HF	.38	.41	.31	.40	.25	.38	.39	.39	.44	.44	.47	.34	.25	.50	.35

APPENDIX F: Raw Score to Percentile Comversion Tables

Table F-1. AFOQT Form P1 Pilot Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-38	01	102	33	136	69
39-44	02	103	34	137	70
45-51	03	104	35	138	71
52-55	04	105	36	139	73
56-57	05	106	37	140	74
58-61	06	107	38	141	75
62-64	07	108	39	142	76
65-67	08	109	41	143	77
68	09	110-111	42	144	78
69-70	10	112	43	145	79
71-72	11	113	44	146	80
73-74	12	114	45	147	81
75-77	13	115	46	148	82
78	14	116	47	149	83
79	15	117	48	150-151	84
80	16	118	50	152	85
81-82	17	119	51	153-154	86
83	18	120	52	155	87
84	19	121	53	156	88
85-87	20	122	54	157	89
88	21	123	55	158	90
89	22	124	56	159	91
90	23	125	57	160	92
91-92	24	126	58	161	93
93	25	127	60	162-163	94
94	26	128	61	164-165	95
95	27	129	62	166-168	96
96-97	28	130-131	63	169-173	97
98	29	132	64	174-179	98
99	30	133	65	180-204	99
100	31	134	66		
101	32	135	67		

Table F-2. AFOQT Form P1 Navigator-Technical Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-58	01	138	34	178	67
59-69	02	139	35	179	68
70-75	03	140-141	36	180	69
76-79	04	142	37	181	70
80-83	05	143-144	38	182	71
84-85	06	145	39	183	72
86-88	07	146	40	184-185	73
89-91	08	147	41	186-187	74
92-95	09	148	42	188	75
96-97	10	149-151	43	189	76
98-99	11	152	44	190	77
100-102	12	153	45	191	78
103-104	13	154	46	192-193	79
105-106	14	155	47	194	80
107-109	15	156	48	195-196	81
110-111	16	157	49	197	82
112-113	17	158	50	198-199	83
114-116	18	159	51	200	84
117	19	160-161	52	201	85
118-119	20	162	53	202-203	86
120-121	21	163	54	204-205	87
122	22	164	55	206-207	88
123-124	23	165	56	208-209	89
125	24	166	57	210-211	90
126-127	25	167	58	212-213	91
128	26	168	59	214	92
129	27	169	60	215-216	93
130	28	170	61	217-219	94
131-132	29	171	62	220-222	95
133-134	30	172-173	63	223-227	96
135	31	174	64	228-232	97
136	32	175-176	65	233-236	98
137	33	177	66	237-264	99

Table F-3. AFOQT Form P1 Academic Aptitude Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-27	01	82	29	112	69
28-34	02	83	31	113	70
35-39	03	84	33	114	71
40-41	04	85	34	115	72
42-45	05	86	35	116	75
46-48	06	87	36	117	76
49-50	07	88	37	118	78
51	08	89-90	38	119	79
52-55	09	91	40	120	80
56-57	10	92	41	121	81
58-59	11	93	43	122	82
60	12	94	44	123	83
61	13	95	45	124	84
62	14	96	47	125	85
63	15	97	49	126	86
64-66	16	98	50	127	87
67	17	99	51	128	88
68-69	18	100	52	129	89
70	19	101	53	130	90
71	20	102-103	54	131	91
72-73	21	104	57	132	92
74	22	105	59	133-134	93
75	23	106	61	135	94
76	24	107	62	136-137	95
77	25	108	63	138-139	96
78	26	109	65	140-141	97
79	27	110	67	142-143	98
80-81	28	111	68	144-150	99

Table F-4. AFOQT Form P1 Verbal Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-14	01	34	23	53	62
15	02	35	24	54	64
16-17	03	36	26	55	67
18	04	37	27	56	69
19	05	38	30	57	72
20	06	39	32	58	74
21	07	40	33	59	77
22	08	41	36	60	78
23	09	42	38	61	81
24	10	43	40	62	84
25	11	44	41	63	86
26	12	45	44	64	87
27	13	46	46	65	90
28	14	47	48	66	92
29	15	48	50	67	93
30	17	49	53	68	96
31	18	50	55	69	97
32	19	51	57	70	98
33	21	52	60	71-75	99

Table F-5. AFOQT Form P1 Quantitative Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-14	01	39	28	60	71
15-16	02	40-41	31	61	75
17-19	03	42	33	62	76
20	04	43-44	34	63	78
21	05	45	38	64	80
22	06	46	41	65	82
23-24	08	47-48	43	66	85
25	09	49	45	67	86
26	10	50	48	68	90
27-28	11	51-52	52	69	91
29	14	53	54	70	92
30	15	54	57	71	94
31-32	17	55	59	72	95
33	19	56	61	73	97
34-35	21	57	64	74	98
36	24	58	66	75	99
37-38	26	59	69		

Table F-6. AFOQT Form P2 Pilot Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-41	01	101	33	134	69
42-46	02	102	34	135	70
47-53	03	103	35	136	71
54-57	04	104	36	137	73
58-59	05	105	37	138	74
60-62	06	106	38	139	75
63-65	07	107	39	140	76
66-67	08	108	41	141	77
68	09	109	42	142-143	78
69-70	10	110	43	144	79
71-73	11	111	44	145	80
74-75	12	112	45	146	81
76-77	13	113	46	147	82
78	14	114	47	148	83
79	15	115	48	149-150	84
80	16	116	50	151	85
81-82	17	117	51	152-153	86
83	18	118	52	154	87
84	19	119-120	53	155	88
85-86	20	121	54	156	89
87	21	122	55	157	90
88	22	123	56	158	91
89	23	124	57	159	92
90-91	24	125	58	160	93
92	25	126	60	161-162	94
93	26	127	61	163-164	95
94	27	128	62	165-168	96
95-96	28	129	63	169-173	97
97	29	130	64	174-179	98
98	30	131	65	180-204	99
99	31	132	66		
100	32	133	67		

Table F-7. AFOQT Form P2 Navigator-Technical Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-59	01	135	34	174	67
60-69	02	136	35	175	68
70-74	03	137	36	176	69
75-79	04	138	37	177	70
80-82	05	139-140	38	178	71
83-84	06	141-142	39	179	72
85-87	07	143	40	180-181	73
88-90	08	144	41	182	74
91-93	09	145	42	183-184	75
94-95	10	146-147	43	185	76
96-98	11	148	44	186	77
99-100	12	149	45	187	78
101-102	13	150	46	188-189	79
103-104	14	151	47	190	80
105-106	15	152	48	191-192	81
107-109	16	153	49	193	82
110-111	17	154-155	50	194-195	83
112-113	18	156	51	196	84
114	19	157	52	197	85
115-116	20	158	53	198	86
117-119	21	159	54	199-201	87
120	22	160	55	202-203	88
121	23	161	56	204-205	89
122	24	162	57	206-207	90
123-124	25	163	58	208-209	91
125	26	164	59	210	92
126	27	165	60	211-212	93
127	28	166	61	213-215	94
128	29	167	62	216-219	95
129-131	30	168-169	63	220-223	96
132	31	170	64	224-228	97
133	32	171-172	65	229-232	98
134	33	173	66	233-264	99

Table F-8. AFOQT Form P2 Academic Aptitude Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-29	01	82	29	114	69
30-34	02	83	31	115	70
35-39	03	84	33	116-117	71
40-41	04	85-86	34	118	72
42-44	05	87	35	119	75
45-47	06	88	36	120	76
48-49	07	89	37	121	78
50	08	90-91	38	122	79
51-54	09	92	40	123	80
55-56	10	93	41	124	81
57	11	94	43	125	82
58-59	12	95	44	126	83
60	13	96-97	45	127	85
61	14	98	47	128	86
62	15	99	49	129	87
63-65	16	100	50	130	88
66	17	101	51	131	89
67-68	18	102	52	132	90
69-70	19	103	53	133	91
71	20	104-105	54	134	92
72	21	106	57	135-136	93
73	22	107	59	137	94
74-75	23	108	61	138	95
76	24	109	62	139-140	96
77	25	110	63	141-142	97
78	26	111	65	143	98
79-80	27	112	67	144-150	99
81	28	113	68		

Table F-9. AFOQT Form P2 Verbal Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-15	01	36	23	55	62
16	02	37	24	56	64
17-18	03	38	26	57	67
19	04	39	27	58	69
20	05	40	30	59	72
21	06	41	32	60	74
22	07	42	33	61	77
23	08	43	36	62	78
24	09	44	38	63	81
25	10	45	40	64	84
26	11	46	41	65	86
27	12	47	44	66	87
28	13	48	46	67	90
29	14	49	48	68	92
30	15	50	50	69	93
31	17	51	53	70	96
32-33	18	52	55	71	97
34	19	53	57	72	98
35	21	54	60	73-75	99

Table F-10. AFOQT Form P2 Quantitative Composite Conversion Table

Raw score	Percentile	Raw score	Percentile	Raw score	Percentile
0-14	01	37-38	28	59	71
15-16	02	39	31	60-61	75
17-18	03	40-41	33	62	76
19	04	42	34	63	78
20	05	43-44	38	64	82
21	06	45	41	65	85
22-23	08	46	43	66	86
24	09	47-48	45	67	88
25	10	49	48	68	90
26-27	11	50	52	69	91
28	14	51	54	70	92
29	15	52-53	57	71	94
30-31	17	54	59	72	95
32	19	55	61	73	96
33	21	56	64	74	98
34-35	24	57	66	75	99
36	26	58	69		

APPENDIX G: AFOQT Bibliography

AFOQT Bibliography

1950-1959

Byers, W.E. (1954). The degree to which the Air Force Officer Qualifying Test Battery measures factors that are independent of more easily obtained data. Unpublished master's thesis, University of Wisconsin.

Flyer, E.S. (1954). AFOQT as a predictor of grades in the aircraft controllers course (PRL-TM-54-5). Lackland AFB, TX: Personnel Research Laboratory.

Roy, H., Brueckel, J.S., & Drucker, A.J. (1954). Selection of Army and Air Force Reserve Officer Training Corps students (Personnel Research Note No. 28). Washington, DC: The Adjutant General's Office.

Staff. (1954). Selection of Army and Air Force Reserve Officer Training Corps students (TRN-28). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Flyer, E.S., & Bigbee, L.R. (1955a). Estimates of the validity of the AFOQT pilot aptitude score for an AFROTC sample (PRL-TM-55-1). Lackland AFB, TX: Personnel Research Laboratory.

Flyer, E.S., & Bigbee, L.R. (1955b). Light plane proficiency ratings as a selection device for AFROTC pilot trainees (PRL-TM-55-38). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Christal, R.E., & Goebel, L.G. (1956). Composition and intercorrelations of the 1955 Air Force Officer Qualifying Test composites (PRL-TN-56-4). Lackland AFB, TX: Personnel Research Laboratory.

Goebel, L.D., & Christal, R.E. (1956). Composition and intercorrelations of the 1956 Air Force Officer Qualifying Test (Form C) composites (PRL-TN-56-6). Lackland AFB, TX: Personnel Research Laboratory.

Hagen, E.P., & Thorndike, R.L. (1956). Follow-up study of Air Force cadet examinees (AFPTRC-TN-56-58, AD-105 553). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Christal, R.E., & Krumboltz, J.D. (1957a). Prediction of first semester criteria at the Air Force Academy (AFPTRC-TN-57-17, AD-098 920). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Christal, R.E., & Krumboltz, J.D. (1957b). Use of the Air Force Officer Qualifying Test in the AFROTC selection program (PL-TM-57-6). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Farr, H.M. (1957). Procurement of Air Force officers during the period 1957 through 1977 (M-35582-5-NC, F239p). Maxwell AFB, AL: Air University, Air Command and Staff College.

Mahoney, H. (1957). An analysis of current personnel selection for navigator training in the USAF (M-32983-NC, M216a). Maxwell AFB, AL: Air University, Air War College.

Tupes, E.C. (1957a). Officer aptitude and officer effectiveness report differences among groups of applicants for regular commission (PL-TM-57-5). Lackland AFB, TX: Personnel Laboratory.

Tupes, E.C. (1957b). Relationships between measures obtained during AFROTC training and later officer effectiveness reports (PL-TM-57-4). Lackland AFB, TX: Personnel Laboratory.

Tupes, E.C., Carp, A., & Borg, W.R. (1957). Validation of a proposed officer effectiveness selection battery (AFPTRC-TN-57-141, AD-146 415). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Tupes, E.C., & Christal, R.E. (1957). Psychological tests and the selection and classification of Air Force officers (AFPTRC-TN-57-52, AD-126 383). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Zaccaria, M.A., Page, W.L., & Ailsworth, K.A. (1957). Predictors of final grade in communications officer course (PL-TM-57-11). Lackland AFB, TX: Personnel Laboratory.

Barron, F., Block, J., MacKinnon, D.W., & Woodworth, D.G. (1958). An assessment study of Air Force officers: Part III. Assessment correlates of criteria of officer effectiveness (WADC-TR-58-91(III), AD-210 218). Lackland AFB, TX: Personnel Laboratory.

Gough, H.G. (1958). An assessment study of Air Force officers: Part IV. Predictability of a composite criterion of officer effectiveness (WADC-TR-58-91(IV), AD-210 219). Lackland AFB, TX: Personnel Laboratory.

Gough, H.G., & Krauss, I. (1958). An assessment study of Air Force officers: Part II. Description of the assessed sample (WADC-TR-58-91 (II), AD-208 700). Lackland AFB, TX: Personnel Laboratory.

MacKinnon, D.W. (1958). An assessment study of Air Force officers: Part V. Summary and applications (WADC-TR-58-91(V), AD-210 220). Lackland AFB, TX: Personnel Laboratory.

MacKinnon, D.W., Crutchfield, R.S., Barron, F., Block, J., Gough, H.G., & Harris, R.E. (1958). An assessment study of Air Force officers: Part I: Design of the study and description of the variables (WADC-TR-58-91 (I), AD-151 040). Lackland AFB, TX: Personnel Laboratory.

Thorndike, R.L., & Hagen, E.P. (1958). Long-term prediction of some officer-effectiveness measures from aptitude tests (WADC-TR-58-489, AD-204 531). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Center.

Valentine, L.D., Jr. (1958). Validity of the AFOQT (Form A) for prediction of student-officer success in observer training (WADC-TN- 58-69, AD-207 334). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Center.

Creager, J.A. (1959). Standardization of Form E, AFOQT (WCLL-TM- 59-12). Lackland AFB, TX: Personnel Laboratory.

Mullins, C.J. (1959). Mean and percentage distributions of FIP trainees on AFOQT scores (WCLL-TM-59-39). Lackland AFB, TX: Personnel Laboratory.

Tupes, E.C. (1959). A comparison of the importance ratings of Air Force officer duty assignments with the average officer aptitude and officer effectiveness of incumbents (WCLL-TM-59-16). Lackland AFB, TX: Personnel Laboratory.

1960-1969

Miller, R.E. (1960a). Comparison of AFOQT composites and CEEB scores in the Air Force Academy selection battery (WWRDR-TM-60-25). Lackland AFB, TX: Wright Air Development Division, Air Research and Development Command.

Miller, R.E. (1960b). Prediction of Officer Training School criteria from the Air Force Officer Qualifying Test (WCLL-TM-60-10). Lackland AFB, TX: Personnel Laboratory.

Miller, R.E. (1960c). Prediction of technical training criteria from AFOQT composites (WADD-TN-60-215, AD-246 658). Lackland AFB, TX: Wright Air Development Division.

Miller, R.E. (1960d). The Cadet Screening Test as a predictor of AFOQT aptitude composites (WCLL-TM-60-14). Lackland AFB, TX: Personnel Laboratory.

Tupes, E.C. (1960). Estimating AFOQT officer quality stanines from general aptitude index (WWRDP-TM-60-15). Lackland AFB, TX: Personnel Laboratory.

Valentine, L.D., Jr. (1960a). Levels of aptitudes of the first three Officer Training School classes (WWRDP-TM-60-16). Lackland AFB, TX: Personnel Laboratory.

Valentine, L.D., Jr. (1960b). Scoring AFOQT booklet I from mark-sense cards (WCLL-TM-60-6). Lackland AFB, TX: Personnel Laboratory.

Creager, J.A., & Miller, R.E. (1961). Summary of regression analyses in the prediction of leadership criteria: Air Force Academy classes of 1961 through 1963 (ASD-TN-61-41, AD-263 979). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Valentine, L.D., Jr., & Creager, J.A. (1961). Officer selection and classification tests: Their development and use (ASD-TN-61-145, AD-269 827). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Miller, R.E., & Valentine, L.D., Jr. (1964). Development and standardization of the Air Force Officer Qualifying Test - 64 (PRL-TDR-64-6, AD-600 782). Lackland AFB, TX: Personnel Research Laboratory.

Miller, R.E. (1966a). Development of officer selection and classification tests - 1966 (PRL-TR-66-5, AD-639 237). Lackland AFB, TX: Personnel Research Laboratory.

Miller, R.E. (1966b). Relationship of AFOQT scores to measures of success in undergraduate pilot and navigator training (PRL-TR-66-14, AD-656 303). Lackland AFB, TX: Personnel Research Laboratory.

Karst, C.F. (1967). USAF regular officer procurement and selection (M-35582-7-U, K18u). Maxwell AFB, AL: Air University, Air Command and Staff College.

Koepnick, R.E. (1967). The selection and training of procurement officers by the United States Air Force (M-35582-7-U, K78s). Maxwell AFB, AL: Air University, Air Command and Staff College.

Shaffer, W.M. (1967). An analysis of current USAF regular commissioning procedures (M-35582-7-U, S5252a). Maxwell AFB, AL: Air University, Air Command and Staff College

Gregg, G. (1968). The effect of maturation and educational experience on Air Force Officer Qualifying Test scores (AFHRL-TR-68-107, AD-687 089). Lackland AFB, TX: Air Force Human Resources Laboratory.

Miller, R.E. (1968). Development of officer selection and classification tests (AFHRL-TR-68-104, AD-679 989). Lackland AFB, TX: Air Force Human Resources Laboratory.

Miller, R.E. (1969). Interpretation and utilization of scores on the AFOQT (AFHRL-TR-69-103, AD-691 001). Lackland AFB, TX: Air Force Human Resources Laboratory.

Tupes, E.C., & Miller, R.E. (1969). Equivalence of AFOQT scores for different educational levels (AFHRL-TR-69-19, AD-703 727). Lackland AFB, TX: Air Force Human Resources Laboratory.

1970-1979

Miller, R.E. (1970). Development and standardization of the Air Force Officer Qualifying Test Form K (AFHRL-TR-70-21, AD-710 602). Lackland AFB, TX: Air Force Human Resources Laboratory.

Black, D.H. (1972). United States Air Force personnel selection (M-35582-7-U, B6274u). Maxwell AFB, AL: Air University, Air Command and Staff College.

Miller, R.E. (1972). Development and standardization of the Air Force Officer Qualifying Test Form L (AFHRL-TR-72-47, AD-754 849). Lackland AFB, TX: Air Force Human Resources Laboratory.

Miller, R.E. (1974). Development and standardization of the Air Force Officer Qualifying Test Form M (AFHRL-TR-74-16, AD-778 837). Lackland AFB, TX: Air Force Human Resources Laboratory.

Shull, W.B. (1975a). An investigation of the Air Force Officer Qualifying Test. Dissertations Abstract International, 36, 4942A.

Shull, W.B. (1975b). An investigation of the Air Force Officer Qualifying Test. Proceedings of the 17th Annual Military Testing Association Conference (pp. 716-731). Indianapolis, IN.

Ree, M.J. (1976). Effects of item-option weighting on the reliability and validity of the AFOQT for pilot selection (AFHRL-TR-76-76, AD-A035 732). Lackland AFB, TX: Air Force Human Resources Laboratory.

Alley, W.E., & Gibson, T.A. (1977). Predicting success in the AFROTC scholarship program (AFHRL-TR-77-11, AD-A041 137). Lackland AFB, TX: Air Force Human Resources Laboratory.

Jackson, D.K., & Gordon, M., Jr. (1977). Development of a weighted selection system for the AFROTC professional officer course. Proceedings of the 19th Annual Military Testing Association Conference (pp. 102-111). San Antonio, TX.

Valentine, L.D., Jr. (1977). Navigator-observer selection research: Development of a new Air Force Officer Qualifying Test navigator-technical composite (AFHRL-TR-77-36, AD-A042 689). Brooks AFB, TX: Air Force Human Resources Laboratory.

Gould, R.B. (1978). Air Force Officer Qualifying Test Form N: Development and standardization (AFHRL-TR-78-43, AD-A059 746). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hunter, D.R., & Thompson, N.A. (1978a). Pilot selection system development (AFHRL-TR-78-33, AD-A058 418). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hunter, D.R., & Thompson, N.A. (1978b). Pilot selection system development. Catalog of Selected Documents in Psychology, 8, 102.

Jackson, D.K., & Gordon, M., Jr. (1978). Weighted selection system for AFROTC applicants--perspective after second year of use. Proceedings of the 20th Annual Military Testing Association Conference (pp. 566-572). Oklahoma City, OK.

1980 - to Present

Albert, W.G. (1980). Computerized algorithms: Evaluation of capability to predict graduation from Air Force training (AFHRL-TR-80-6, AD-A091 105). Brooks AFB, TX: Air Force Human Resources Laboratory.

Roach, B.W. (1983). Monetary value of pilot selection using the AFOQT. Proceedings of the 25th Annual Military Testing Association Conference (pp. 478-483). Gulf Shores, AL.

Roach, B.W., & Rogers, D.L. (1983). Sex as a moderator variable in predicting OTS attrition. Proceedings of the 25th Annual Military Testing Association Conference (pp. 496-501). Gulf Shores, AL.

Rogers, D.L. (1983). Development of the Air Force officer screening composites. Proceedings of the 25th Annual Military Testing Association Conference (pp. 467-471). Gulf Shores, AL.

Arth, T.O. (1984). Validation of the Air Force Officer Qualifying Test. Proceedings of the 26th Annual Military Testing Association Conference (pp. 1055-1059). Munich, West Germany.

Skinner, M.J. (1984). Aptitude selectors for minuteman missile combat crew duty. Proceedings of the 26th Annual Military Testing Association Conference (pp. 1073-1079). Munich, West Germany.

Arth, T.O. (1985, August). Which Air Force candidates benefit most from retesting? Paper presented at the 93rd Annual Conference of the American Psychological Association, Los Angeles, CA.

Finegold, L.S., & Rogers, D. (1985). Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training (AFHRL-TR-85-13, AD-A158 162). Brooks AFB, TX: Air Force Human Resources Laboratory.

Rogers, D.L. (1985). Screening composites for Air Force officers (AFHRL-TP-85-2, AD-A154 315). Brooks AFB, TX: Air Force Human Resources Laboratory.

Arth, T.O. (1986a). Air Force Officer Qualifying Test (AFOQT): Retesting effects (AFHRL-TP-86-8, AD-A168 926). Brooks AFB, TX: Air Force Human Resources Laboratory.

Arth, T.O. (1986b). Validation of the AFOQT for nonrated officers (AFHRL-TP-85-50, AD-A164 134). Brooks AFB, TX: Air Force Human Resources Laboratory.

Arth, T.O., & Skinner, M.J. (1986, November). Aptitude selectors for Air Force officers non-aircrew jobs. Proceedings of the 28th Annual Military Testing Association Conference (pp. 301-306). New London, CT.

Rogers, D.L., Roach, B.W., & Short, L.O. (1986). Mental ability testing in the selection of Air Force officers: A brief historical overview (AFHRL-TP-86-23, AD-A173 484). Brooks AFB, TX: Air Force Human Resources Laboratory.

Rogers, D.L., Roach, B.W., & Wegner, T.G. (1986). Air Force Officer Qualifying Test Form O: Development and standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Air Force Human Resources Laboratory.

Wegner, T.G., & Short, L.O. (1986, November). Assessing the accuracy of the AFOQT quick score procedure. Proceedings of the 28th Annual Military Testing Association Conference (pp. 60-65). New London, CT.

Skinner, M.J., & Ree, M.J. (1987). Air Force Officer Qualfying Test (AFOQT): Item and factor analysis of Form O (AFHRL-TR-86-68, AD-A184 975). Brooks AFB, TX: Air Force Human Resources Laboratory.

Berger, F.R., Gupta, W.B., Berger, R.M., & Skinner, J. (1988). Air Force Officer Qualifying Test (AFOQT) Form P: Test construction (AFHRL-TR-88-30, AD-A200 678). Brooks AFB, TX: Air Force Human Resources Laboratory.

Cowan, D.K., & Sperl, T.C. (1988). Selection and classification of United States military officers: A fifty-year bibliography (1937-1986) (AFHRL-TP-88-45, AD-A205 631). Brooks AFB, TX: Air Force Human Resources Laboratory.

Elliott, L.R. (1988). Validation of the Air Force Reserve Officer Training Corps selection system. Proceedings of the 30th Annual Military Testing Association Conference (pp. 469-474). Arlington, VA.

Sperl, T.C. (1988). The development of Quick Score Composites for the Air Force Officer Qualifying Test. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Steuck, K.W., Watson, T.W., & Skinner, J. (1988). Air Force Officer Qualifying Test (AFOQT): Forms P pre-implementation analyses and equating (AFHRL-TP-88-6, AD-A201 100). Brooks AFB, TX: Air Force Human Resources Laboratory.

Barrett, L.E. (1989). Cognitive versus noncognitive predictors of Air Force officer training. Paper presented at the 97th Annual Conference of the American Psychological Association, New Orleans, LA.

Gupta, W.B., Berger, F.R., Berger, R.M., & Skinner, J. (1989). Air Force Officer Qualifying Test (AFOQT): Development of an item bank (AFHRL-TR-89-33, AD-A216 228). Brooks AFB, TX: Air Force Human Resources Laboratory.

Sperl, T.C., & Ree, M.J. (1989). Air Force Officer Qualifying Test (AFOQT): Development of Quick Score Composites for Forms P1 and P2. Paper presented at the 97th Annual Conference of the American Psychological Association, New Orleans, LA.

U.S. GOVERNMENT PRINTING OFFICE